S1 Proofs

We provide proofs of the results, including

- Proposition 2 (Section S1.1),
- Theorem 1 and corollaries (Section S1.2),
- Theorem 2 and corollaries (Section S1.3),
- Propositions 3, 4, 6, and 8 (Section S1.4),
- Proposition 5 (Section S1.5),
- Theorem 3 (Section S1.6),
- Theorem 4 (Section S1.7), and
- Theorem 5 (Section S1.8).

S1.1 Proof of Proposition 2

Without loss of generality, assume that $\ell = 1$ and $0 \leq X_{1}^{(1)} < \ldots < X_{n}^{(1)} \leq 1$. The spline result for $m \geq 1$ follows easily from Proposition 1 in Mammen & van de Geer (1997).

Consider the case $m = 1$. For any $g = \sum_{j=1}^{p} g_{j}$ with $g_{1} \in \mathcal{V}^{1}$, define $\tilde{g}_{1}$ as a piecewise constant function: $\tilde{g}_{1}(z) = g_{1}(X_{i}^{(1)})$ for $0 \leq z < X_{2}^{(1)}$, $\tilde{g}_{1}(z) = g_{1}(X_{i}^{(1)})$ for $X_{1}^{(1)} \leq z < X_{i+1}^{(1)}$, $i = 2, \ldots, n - 1$, and $\tilde{g}_{1}(z) = g_{1}(X_{n}^{(1)})$ for $X_{n}^{(1)} \leq z \leq 1$. Let $\tilde{g} = \tilde{g}_{1} + \sum_{j=2}^{p} g_{j}$. Then $\tilde{g}(X_{i}) = g(X_{i})$ for $i = 1, \ldots, n$, but $\text{TV}(\tilde{g}_{1}) \leq \text{TV}(g_{1})$ and hence $R_{n}(\tilde{g}) \leq R_{n}(g)$, which implies the desired result for $m = 1$.

Consider the case $m = 2$. For any $g = \sum_{j=1}^{p} g_{j}$ with $g_{1} \in \mathcal{V}^{2}$, define $\tilde{g}_{1}$ such that $\tilde{g}_{1}(X_{i}^{(1)}) = g_{1}(X_{i}^{(1)})$, $i = 1, \ldots, n$, and $\tilde{g}_{1}(z)$ is linear in the intervals $[0, X_{2}^{(1)}]$, $[X_{i}^{(1)}, X_{i+1}^{(1)}]$, $i = 2, \ldots, n - 2$, and $[X_{n-1}^{(1)}, 1]$. Then $\text{TV}(\tilde{g}_{1}^{(1)}) = \sum_{i=1}^{n-1} |b_{i+1} - b_{i}|$, where $b_{i}$ is the slope of $\tilde{g}_{1}$ between
On the other hand, by the mean-value theorem, there exists \( z_i \in [X_i^{(1)}, X_{i+1}^{(1)}] \) such that \( g_i^{(1)}(z_i) = b_i \) for \( i = 1, \ldots, n - 1 \). Then \( TV(g_i^{(1)}) \geq \sum_{i=1}^{n-1} |b_{i+1} - b_i| \). Let \( \tilde{g} = \tilde{g}_1 + \sum_{j=2}^p g_j \). Then \( g(X_i) = g(X_i) \) for \( i = 1, \ldots, n \), but \( R_n(\tilde{g}) \leq R_n(g) \), which implies the desired result for \( m = 2 \).

### S1.2 Proofs of Theorem 1 and corollaries

We split the proof of Theorem 1 and Corollary 1 into five lemmas. The first one provides a probability inequality controlling the magnitude of \( \langle \varepsilon, h_j \rangle_n \), in terms of the semi-norm \( \|h_j\|_{F,j} \) and the norm \( \|h_j\|_n \) for all \( h_j \in G_j \) with a single \( j \).

#### Lemma 1.

For fixed \( j \in \{1, \ldots, p\} \), let

\[
A_{nj} = \bigcup_{h_j \in G_j} \{ \langle \varepsilon, h_j \rangle_n / C_1 > \gamma_{nj,t} w_{nj} \|h_j\|_{F,j} + \gamma_{nj,t} \|h_j\|_n \},
\]

where \( \gamma_{nj,t} = \gamma_{nj} + \sqrt{t/n} \) for \( t > 0 \), \( \gamma_{nj} = n^{-1/2} \psi_{nj}(w_{nj}) / w_{nj} \), and \( w_{nj} \in (0, 1] \). Under Assumptions 1 and 2, we have

\[
P(A_{nj}) \leq \exp(-t).
\]

**Proof.** In the event \( A_{nj} \), we renormalize \( h_j \) by letting \( f_j = h_j / (\|h_j\|_{F,j} + \|h_j\|_n / w_{nj}) \). Then \( \|f_j\|_{F,j} + \|f_j\|_n / w_{nj} = 1 \) and hence \( f_j \in G_j(w_{nj}) \). By Lemma 12 with \( F_1 = G_j(w_{nj}) \) and \( \delta = w_{nj} \), we have for \( t > 0 \),

\[
P(A_{nj}) \leq P \left\{ \sup_{f_j \in G_j(w_{nj})} |\langle \varepsilon, f_j \rangle_n| / C_1 > \gamma_{nj,t} w_{nj} \right\}
\]

\[
= P \left\{ \sup_{f_j \in G_j(w_{nj})} |\langle \varepsilon, f_j \rangle_n| / C_1 > n^{-1/2} \psi_{nj}(w_{nj}) + w_{nj} \sqrt{t/n} \right\} \leq \exp(-t).
\]

By Lemma 1 and the union bound, we obtain a probability inequality controlling the magnitude of \( \langle \varepsilon, h_j \rangle_n \) for \( h_j \in G_j \) simultaneously over \( j = 1, \ldots, p \).

#### Lemma 2.

For each \( j \in \{1, \ldots, p\} \), let

\[
A_{nj} = \bigcup_{h_j \in G_j} \{ |\langle \varepsilon, h_j \rangle_n| > \lambda_{nj} w_{nj} \|h_j\|_{F,j} + \lambda_{nj} \|h_j\|_n \},
\]

where \( \lambda_{nj} / C_1 = \gamma_{nj} + \{\log(p/\epsilon) / n\}^{1/2} \). Under Assumptions 1 and 2, we have

\[
P(\bigcup_{j=1}^p A_{nj}) \leq \epsilon.
\]
Proof. By Lemma 1 with \(t = \log(p/\epsilon)\), we have for \(j = 1, \ldots, p\),
\[
P(A_{nj}) \leq \exp(-t) = \frac{\epsilon}{p}.
\]
Applying the union bound yields the desired inequality. \(\square\)

If \(g^* \in \mathcal{G}\), then \(K_n(\hat{g}) \leq K_n(g^*)\) directly gives the basic inequality:
\[
\frac{1}{2}\|\hat{g} - g^*\|_n^2 + A_0 R_n(\hat{g}) \leq \langle \varepsilon, \hat{g} - g^* \rangle_n + A_0 R_n(g^*).
\] (S1)

By exploiting the convexity of the regularizer \(R_n(\cdot)\), we provide a refinement of the basic inequality (S1), which relates the estimation error of \(\hat{g}\) to that of any additive function \(\bar{g} \in \mathcal{G}\) and the corresponding regularization \(R_n(\bar{g})\).

Lemma 3. The fact that \(\hat{g}\) is a minimizer of \(K_n(g)\) implies that for any function \(\bar{g}(x) = \sum_{j=1}^p \bar{g}_j(x^{(j)}) \in \mathcal{G}\),
\[
\frac{1}{2}\|\hat{g} - g^*\|_n^2 + \frac{1}{2}\|\bar{g} - \bar{g}\|_n^2 + A_0 R_n(\hat{g})
\leq \frac{1}{2}\|\bar{g} - g^*\|_n^2 + \langle \varepsilon, \bar{g} - \bar{g}\rangle_n + A_0 R_n(\bar{g}).
\] (S2)

Proof. For any \(t \in (0, 1]\), the fact that \(K_n(\hat{g}) \leq K_n((1-t)\hat{g} + \bar{g})\) implies
\[
\frac{t^2}{2}\|\hat{g} - \bar{g}\|_n^2 + R_n(\hat{g}) \leq \langle Y - ((1-t)\hat{g} + \bar{g}), t(\hat{g} - \bar{g})\rangle_n + R_n((1-t)\hat{g} + \bar{g})
\leq \langle Y - ((1-t)\hat{g} + \bar{g}), t(\hat{g} - \bar{g})\rangle_n + (1-t)R_n(\hat{g}) + tR_n(\bar{g}),
\]
by similar calculation leading to the basic inequality (S1) and by the convexity of \(R_n(\cdot)\): \(R_n((1-t)\hat{g} + \bar{g}) \leq (1-t)R_n(\hat{g}) + tR_n(\bar{g})\). Using \(Y = g^* + \varepsilon\), simple manipulation of the preceding inequality shows that for any \(t \in (0, 1]\),
\[
\langle \hat{g} - g^*, \hat{g} - \bar{g}\rangle_n - \frac{t}{2}\|\hat{g} - \bar{g}\|_n^2 + R_n(\hat{g}) \leq \langle \varepsilon, \hat{g} - \bar{g}\rangle_n + R_n(\bar{g}),
\]
which reduces to
\[
\frac{1}{2}\|\hat{g} - g^*\|_n^2 + \frac{1-t}{2}\|\hat{g} - \bar{g}\|_n^2 + R_n(\hat{g}) \leq \frac{1}{2}\|\bar{g} - g^*\|_n^2 + \langle \varepsilon, \bar{g} - \bar{g}\rangle_n + R_n(\bar{g})
\]
by the fact that \(2(\hat{g} - g^*, \hat{g} - \bar{g})_n = \|\hat{g} - g^*\|_n^2 + \|\hat{g} - \bar{g}\|_n^2 - \|\bar{g} - g^*\|_n^2\). Letting \(t \searrow 0\) yields the desired inequality (S2). \(\square\)

From Lemma 3, we obtain an upper bound of the estimation error of \(\hat{g}\) when the magnitudes of \(\langle \varepsilon, \hat{g}_j - \bar{g}_j\rangle_n, j = 1, \ldots, p\), are controlled by Lemma 2.
Lemma 4. Let \( A_n = \bigcup_{j=1}^p A_{nj} \) with \( h_j = \hat{g}_j - \bar{g}_j \) in Lemma 2. In the event \( A_n^c \), we have for any subset \( S \subset \{1, 2, \ldots, p\} \),

\[
\frac{1}{2} \|\hat{g} - g^*\|_n^2 + \frac{1}{2} \|\bar{g} - \bar{g}\|_n^2 + (A_0 - 1)R_n(\hat{g} - \bar{g}) \leq \Delta_n(\bar{g}, S) + 2A_0 \sum_{j \in S} \lambda_{nj}\|\hat{g}_j - \bar{g}_j\|_n, \tag{S3}
\]

where

\[
\Delta_n(\bar{g}, S) = \frac{1}{2} \|\bar{g} - g^*\|_n^2 + 2A_0 \left( \sum_{j=1}^p \rho_{nj}\|\hat{g}_j\|_{F,j} + \sum_{j \in S^c} \lambda_{nj}\|\hat{g}_j\|_n \right).
\]

Proof. By the refined basic inequality (S2), we have in the event \( A_n^c \),

\[
\frac{1}{2} \|\hat{g} - g^*\|_n^2 + \frac{1}{2} \|\bar{g} - \bar{g}\|_n^2 + A_0 R_n(\hat{g}) \leq \frac{1}{2} \|\bar{g} - g^*\|_n^2 + R_n(\hat{g} - \bar{g}) + A_0 R_n(\bar{g}).
\]

Applying to the preceding inequality the triangle inequalities,

\[
\|\hat{g}_j\|_{F,j} \geq \|\hat{g}_j - \bar{g}_j\|_{F,j} - \|\bar{g}_j\|_{F,j}, \quad j = 1, \ldots, p,
\]

\[
\|\hat{g}_j\|_n \geq \|\hat{g}_j - \bar{g}_j\|_n - \|\bar{g}_j\|_n, \quad j \in S^c,
\]

\[
\|\hat{g}_j\|_n \geq \|\bar{g}_j\|_n - \|\hat{g}_j - \bar{g}_j\|_n, \quad j \in S,
\]

and rearranging the result leads directly to (S3).

Taking \( S = \emptyset \) in (S3) yields (15) in Corollary 1. In general, we derive implications of (S3) by invoking the compatibility condition (Assumption 3).

Lemma 5. Suppose that Assumption 3 holds. If \( A_0 > (\xi_0 + 1)/(\xi_0 - 1) \), then (S3) with \( S \) from Assumption 3 implies (14) in Theorem 1.

Proof. For the subset \( S \) used in Assumption 3, write

\[
Z_n = \frac{1}{2} \|\hat{g} - g^*\|_n^2 + \frac{1}{2} \|\bar{g} - \bar{g}\|_n^2,
\]

\[
T_{n1} = \sum_{j=1}^p \rho_{nj}\|\hat{g}_j - \bar{g}_j\|_{F,j} + \sum_{j \in S^c} \lambda_{nj}\|\hat{g}_j - \bar{g}_j\|_n, \quad T_{n2} = \sum_{j \in S} \lambda_{nj}\|\hat{g}_j - \bar{g}_j\|_n.
\]

Inequality (S3) can be expressed as

\[
Z_n + (A_0 - 1)(T_{n1} + T_{n2}) \leq \Delta_n(\bar{g}, S) + 2A_0 T_{n2},
\]
which leads to two possible cases: either

\[ Z_n + \xi_1(A_0 - 1)(T_{n1} + T_{n2}) \leq \Delta_n(\bar{g}, S), \] (S4)

or \((1 - \xi_1)(A_0 - 1)(T_{n1} + T_{n2}) \leq 2A_0T_n2,\) that is,

\[ (A_0 - 1)(T_{n1} + T_{n2}) \leq \frac{2A_0}{1 - \xi_1}T_{n2} = (\xi_0 + 1)(A_0 - 1)T_{n2}, \] (S5)

where \(\xi_1 = 1 - 2A_0/(\xi_0 + 1)(A_0 - 1)\) \((\xi_0 > 1)/((\xi_0 - 1).\) If (S5) holds, then \(T_{n1} \leq \xi_0T_{n2},\) which, by Assumption 3 with \(f_j = \bar{g}_j - \bar{g}_j,\) implies

\[ T_{n2} \leq \kappa_0^{-1}\left(\sum_{j \in S} \lambda_{nj}^2\right)^{1/2} \|\bar{g} - \bar{g}\|_n. \] (S6)

Substituting (S6) into the right hand side of (S3) and using \(2ab \leq a^2 + b^2\) yields

\[ Z_n + (A_0 - 1)(T_{n1} + T_{n2}) \leq \Delta_n(\bar{g}, S) + 2A_0\kappa_0^{-1}\left(\sum_{j \in S} \lambda_{nj}^2\right)^{1/2} \|\bar{g} - \bar{g}\|_n \]
\[ \leq \Delta_n(\bar{g}, S) + \frac{1 - \xi_1}{2} \|\bar{g} - \bar{g}\|_n^2 + \frac{2A_0^2}{1 - \xi_1}\kappa_0^{-2}\left(\sum_{j \in S} \lambda_{nj}^2\right). \] (S7)

Therefore, inequality (S3), through (S4) and (S7), implies

\[ \frac{1}{2} \|\bar{g} - g^*\|_n^2 + \frac{\xi_1}{2} \|\bar{g} - \bar{g}\|_n^2 + \xi_1(A_0 - 1)(T_{n1} + T_{n2}) \leq \Delta_n(\bar{g}, S) + \xi_2A_0\kappa_0^{-2}\left(\sum_{j \in S} \lambda_{nj}^2\right). \]

\(\square\)

Finally, combining Lemmas 2, 4 and 5 completes the proof of Theorem 1.

**Proof of Corollary 2.** The result follows from upper bounds of \(\sum_{j \in S} \lambda_{nj}^2\) and \(\sum_{j \in S^c} \lambda_{nj}\|\bar{g}_j\|_n\) by the definition \(S = \{1 \leq j \leq p : \|\bar{g}_j\|_n > C_0\lambda_{nj}\}.\) First, because \(\sum_{j=1}^p \lambda_{nj}^{2-q}\|\bar{g}_j\|_n^q \geq \sum_{j \in S} \lambda_{nj}^{2-q}(C_0^+)^{q}\lambda_{nj}^q,\) we have

\[ \sum_{j \in S} \lambda_{nj}^2 \leq (C_0^+)^{-q}\sum_{j=1}^p \lambda_{nj}^{2-q}\|\bar{g}_j\|_n^q, \] (S8)

where for \(z \geq 0, (z^+)^q = z^q\) if \(q > 0\) or \(1 = 1\) if \(q = 0.\) Second, because \(\sum_{j \in S^c} \lambda_{nj}\|\bar{g}_j\|_n \leq \sum_{j=1}^p \lambda_{nj}(C_0\lambda_{nj})^{1-q}\|\bar{g}_j\|_n^q,\) we have

\[ \sum_{j \in S^c} \lambda_{nj}\|\bar{g}_j\|_n \leq C_0^{1-q}\sum_{j=1}^p \Lambda_{nj}^{2-q}\|\bar{g}_j\|_n^q. \] (S9)
Inserting (S8) and (S9) into (14) yields the desired inequality.  

Proof of Corollary 3. The result follows directly from Corollary 2, because  \( \lambda_{nj}^{2-q} = C_{1}^{2-q}(\gamma_{n}(q) + \nu_{n})^{2-q} \) and  \( \rho_{nj} = C_{1}(\gamma_{n}(q) + \nu_{n})\gamma_{1-q}(q) \leq C_{1}(\gamma_{n}(q) + \nu_{n})^{2-q} \), where  \( \nu_{n} = (\log(p/\epsilon)/n)^{1/2} \).  

S1.3 Proofs of Theorem 2 and corollaries

Write  \( h_{j} = \hat{g}_{j} - \bar{g}_{j} \) and  \( h = \hat{g} - \bar{g} \) and, for the subset  \( S \) used in Assumption 5,

\[
Z_{n} = \frac{1}{2} \| \hat{g} - g^{*} \|_{n}^{2} + \frac{1}{2} \| h \|_{n}^{2},
\]

\[
T_{n1}^{*} = \sum_{j=1}^{p} \rho_{nj} \| h_{j} \|_{F,j} + \sum_{j \in S^{c}} \lambda_{nj} \| h_{j} \|_{Q}, \quad T_{n2}^{*} = \sum_{j \in S} \lambda_{nj} \| h_{j} \|_{Q}.
\]

Compared with the definitions in Section S1.2,  \( Z_{n} \) is the same as before, and  \( T_{n1}^{*} \) and  \( T_{n2}^{*} \) are similar to  \( T_{n1} \) and  \( T_{n2} \), but with  \( \| h_{j} \|_{Q} \) used instead of  \( \| h_{j} \|_{n} \).

Let

\[
\Omega_{n1} = \left\{ \sup_{g \in \mathcal{G}} \frac{\| g \|_{n}^{2} - \| g \|_{Q}^{2}}{R^{2}(g)} \leq \phi_{n} \right\}.
\]

Then  \( P(\Omega_{n1}) \geq 1 - \pi \). In the event  \( \Omega_{n1} \), we have by Assumption 6(i),

\[
\max_{j=1,...,p} \sup_{g_{j} \in \mathcal{G}_{j}} \frac{\| g_{j} \|_{n} - \| g_{j} \|_{Q}}{w_{nj} \| g_{j} \|_{F,j} + \| g_{j} \|_{Q}} \leq (\max_{j=1,...,p} \lambda_{nj})^{1/2} \phi_{n} \leq \eta_{0}. \tag{S10}
\]

By direct calculation, (S10) implies that if  \( \| g_{j} \|_{F,j} + \| g_{j} \|_{n}/w_{nj} \leq 1 \) then  \( \| g_{j} \|_{F,j} + \| g_{j} \|_{Q}/w_{nj} \leq (1 - \eta_{0})^{-1} \). Hence (S10) implies that

\[
H(u, \mathcal{G}_{j}(w_{nj}), \| \cdot \|_{n}) \leq H((1 - \eta_{0})u, \mathcal{G}_{j}^{*}(w_{nj}), \| \cdot \|_{n}),
\]

and  \( \psi_{nj}(w_{nj}) \) satisfying (19) also satisfies (13) for  \( \delta = w_{nj} \). Let  \( \Omega_{n2} = A_{n}^{c} \) in Lemma 4. Then conditionally on  \( X_{1:n} = (X_{1}, \ldots, X_{n}) \) for which  \( \Omega_{n1} \) occurs, we have  \( P(\Omega_{n2}|X_{1:n}) \geq 1 - \epsilon \) by Lemma 2. Therefore,  \( P(\Omega_{n1} \cap \Omega_{n2}) \geq (1 - \epsilon)(1 - \pi) \geq 1 - \epsilon - \pi \).

In the event  \( \Omega_{n2} \), recall that (S3) holds, that is,

\[
Z_{n} + (A_{0} - 1)R_{n}(h) \leq \Delta_{n}(\hat{g}, S) + 2A_{0} \sum_{j \in S} \lambda_{nj} \| h_{j} \|_{n}. \tag{S11}
\]

In the event  \( \Omega_{n1} \cap \Omega_{n2} \), simple manipulation of (S11) using (S10) shows that

\[
Z_{n} + A_{1} R_{n}^{*}(h) \leq \Delta_{n}^{*}(\hat{g}, S) + 2A_{0} \sum_{j \in S} \lambda_{nj} \| h_{j} \|_{Q}, \tag{S12}
\]
where \( A_1 = (A_0 - 1) - \eta_0(A_0 + 1) > 0 \) because \( A_0 > (1 + \eta_0)/(1 - \eta_0) \). In the following, we restrict to the event \( \Omega_{n1} \cap \Omega_{n2} \) with probability at least \( 1 - \epsilon - \pi \).

**Proof of Corollary 4.** Taking \( S = \emptyset \) in (S12) yields (28), that is,

\[
Z_n + A_1 R_n^*(h) \leq \Delta_n^*(\bar{g}, \bar{g}^*, \emptyset).
\]

As a result, \( R_n^*(h) \leq A_1^{-1} \Delta_n^*(\bar{g}, \emptyset) \) and hence \( h \|Q \leq \|h\|^2_n + \phi_n R_n^*(h) \leq \|h\|^2_n + \phi_n A_1^{-2} \Delta_n^2(\bar{g}, \emptyset) \). Inequality (29) then follows from (28).

**Proof of Theorem 2.** Inequality (S12) can be expressed as

\[
Z_n + A_1(T_{n1}^* + T_{n2}^*) \leq \Delta_n^*(\bar{g}, S) + 2A_0 T_{n2}^*,
\]

which leads to two possible cases: either

\[
Z_n + A_1(T_{n1}^* + T_{n2}^*) \leq \Delta_n^*(\bar{g}, S), \quad (S13)
\]

or \( (1 - \xi_1)A_1(T_{n1}^* + T_{n2}^*) \leq 2A_0 T_{n2}^* \), that is,

\[
A_1(T_{n1}^* + T_{n2}^*) \leq \frac{2A_0}{1 - \xi_1} T_{n2}^* = (\xi_0^* + 1)A_1 T_{n2}^* = \xi_2 T_{n2}^*, \quad (S14)
\]

where \( \xi_1^* = 1 - 2A_0/\{((\xi_0^* + 1)A_1) \in (0, 1) \) because \( A_0 > \{\xi_0^* + 1 + \eta_0(\xi_0^* + 1)\}/\{\xi_0^* - 1 - \eta_0(\xi_0^* + 1)\} \). If (S14) holds, then \( T_{n1}^* \leq \xi_0^* T_{n2}^* \), which, by the theoretical compatibility condition (Assumption 5) with \( f_j = \bar{g}_j - \tilde{g}_j \), implies

\[
T_{n2}^* \leq \kappa_0^{-1} \left( \sum_{j \in S} \lambda_{nj}^2 \right)^{1/2} \|h\|_Q \quad (S15)
\]

\[
\leq \kappa_0^{-1} \left( \sum_{j \in S} \lambda_{nj}^2 \right)^{1/2} \left\{ \|h\|_n + \phi_n^{1/2}(T_{n1}^* + T_{n2}^*) \right\} \quad (S16)
\]

Substituting (S16) into the right hand side of (S12) and using Assumption 6(ii), \( \phi_n^{1/2}(\xi_0^* + 1) \kappa_0^{-1}(\sum_{j \in S} \lambda_{nj}^2)^{1/2} \leq \eta_1 \), yields

\[
Z_n + A_1(T_{n1}^* + T_{n2}^*) \leq \Delta_n^*(\bar{g}, S) + 2A_0 \kappa_0^{-1} \left( \sum_{j \in S} \lambda_{nj}^2 \right)^{1/2} \left\{ \|h\|_n + \phi_n^{1/2}(T_{n1}^* + T_{n2}^*) \right\}
\]

\[
\leq \Delta_n^*(\bar{g}, S) + 2A_0 \kappa_0^{-1} \left( \sum_{j \in S} \lambda_{nj}^2 \right)^{1/2} \|h\|_n + (1 - \xi_1^*)A_1 \eta_1 (T_{n1}^* + T_{n2}^*). \]
Using \(0 \leq \eta_1 \leq 1\) and \(2ab \leq a^2 + b^2\) in the above inequality leads to

\[
Z_n + \xi_1^* A_1(T_{n1}^* + T_{n2}^*) \leq \Delta_n^*(\tilde{g}, S) + \frac{1 - \xi_1^*}{2} \|h\|^2_n + \frac{2A_0^2}{1 - \xi_1^*} \kappa_0^{-2} \left( \sum_{j \in S} \lambda_{n_j}^2 \right).
\]  

(S17)

Therefore, inequality (S12), through (S13) and (S17), implies (26):

\[
\frac{1}{2} \|\tilde{g} - g^*\|^2_n + \frac{\xi_1^*}{2} \|h\|^2_n + \xi_1^* A_1(T_{n1}^* + T_{n2}^*) \leq \Delta_n^*(\tilde{g}, S) + \xi_1^* A_0 \kappa_0^{-2} \left( \sum_{j \in S} \lambda_{n_j}^2 \right).
\]

To demonstrate (27), we return to the two possible cases, (S13) or (S14). On one hand, if (S13) holds, then \(A_1 R_n^*(h) = A_1(T_{n1}^* + T_{n2}^*)\) is also bounded from above by \(\xi_1^{*-1}\) times the right hand side of (S13) and hence

\[
\|h\|^2_Q \leq \|h\|^2_n + \phi_n R_n^{2*}(h) \leq \|h\|^2_n + \frac{\phi_n}{A_1^2} \xi_1^{*-2} \Delta_n^2(\tilde{g}, S).
\]

(S18)

Simple manipulation of (S13) using (S18) yields

\[
\frac{1}{2} \|\tilde{g} - g^*\|^2_n + \frac{1}{2} \|h\|^2_Q + \xi_1^* A_1(T_{n1}^* + T_{n2}^*) \leq \Delta_n^*(\tilde{g}, S) + \frac{\phi_n}{2A_1^2} \xi_1^{*-2} \Delta_n^2(\tilde{g}, S).
\]

(S19)

On the other hand, combining (S14) and (S15) yields

\[
A_1(T_{n1}^* + T_{n2}^*) \leq \xi_2^* \kappa_0^{-1} \left( \sum_{j \in S} \lambda_{n_j}^2 \right)^{1/2} \|h\|_Q.
\]

(S20)

As a result, \(A_1 R_n^*(h) = A_1(T_{n1}^* + T_{n2}^*)\) is also bounded from above by the right hand side of (S20) and hence by Assumption 6(ii),

\[
\|h\|^2_Q \leq \|h\|^2_n + \phi_n R_n^{2*}(h)
\]

\[
\leq \|h\|^2_n + \frac{\phi_n}{A_1^2} \xi_2^* \kappa_0^{-2} \left( \sum_{j \in S} \lambda_{n_j}^2 \right) \|h\|^2_Q \leq \|h\|^2_n + \frac{1}{2} \eta_1^2 \|h\|^2_Q.
\]

(S21)

Substituting (S15) into the right hand side of (S12) and using (S21) and \(2ab \leq a^2 + b^2\) yields

\[
\frac{1}{2} \|\tilde{g} - g^*\|^2_n + \frac{1 - \eta_1^2}{2} \|h\|^2_Q + A_1(T_{n1}^* + T_{n2}^*) \leq \Delta_n^*(\tilde{g}, S) + 2A_0 \kappa_0^{-1} \left( \sum_{j \in S} \lambda_{n_j}^2 \right)^{1/2} \|h\|_Q
\]

\[
\leq \Delta_n^*(\tilde{g}, S) + \frac{(1 - \eta_1^2)(1 - \xi_1^* )}{2} \|h\|^2_Q + \frac{2A_0^2}{(1 - \eta_1^2)(1 - \xi_1^*) \kappa_0^{-2}} \left( \sum_{j \in S} \lambda_{n_j}^2 \right).
\]

(S22)
Therefore, inequality (S12), through (S19) and (S22), implies (27):

\[ \frac{1}{2} \| \hat{g} - g^\ast \|^2 + \frac{(1 - \eta_2^2)\xi_1^*}{2} \| h \|^2_Q + \xi_1^* A_1 (T_{n1}^* + T_{n2}^*) \]

\[ \leq \Delta_n^* (\hat{g}, S) + \frac{\xi_2^* A_0}{1 - \eta_1^2} k_0^{n^2 - 2} \left( \sum_{j \in S} \lambda_{nj}^2 \right) + \frac{\phi_n}{2A_1^2} \xi_1^* - 2 \Delta_n^2 (\hat{g}, S). \]

**Proof of Corollary 5.** We use the following upper bounds, obtained from (S8) and (S9) with \( S = \{ 1 \leq j \leq p : \| \hat{g}_j \| > C_0^* \lambda_{nj} \} \),

\[ \sum_{j \in S} \lambda_{nj}^2 \leq (C_0^* + \eta) \sum_{j=1}^p \lambda_{nj}^2 \| \hat{g}_j \|^2_Q, \tag{S23} \]

and

\[ \sum_{j \in S^c} \lambda_{nj} \| \hat{g}_j \|_Q \leq C_0^* \sum_{j=1}^p \lambda_{nj}^2 \| \hat{g}_j \|^2_Q. \tag{S24} \]

Equations (25) and (30) together imply \( \phi_n \Delta_n^* (\hat{g}, S) = O(1) + \phi_n \| \hat{g} - g^\ast \|^2/Q. \) Inserting this into (27) and applying (S23) and (S24) yields the high-probability result about \( D_n^* (\hat{g}, \hat{g}). \) The in-probability result follows by using \( \epsilon \to 0, \| \hat{g} - g^\ast \|^2 = O_p(1) \| \hat{g} - g^\ast \|^2/Q, \) by the Markov inequality, and \( \| \hat{g} - g^\ast \|^2 \leq 2 \| \hat{g} - \hat{g} \|^2 + \| \hat{g} - g^\ast \|^2 \) by the triangle inequality. \( \square \)

**Proof of Corollary 6.** First, we show

\[ w_n^* (q) \leq \{ \gamma_n (q) + \nu_n \}^{1-q}. \tag{S25} \]

In fact, if \( \gamma_n (q) \geq \nu_n, \) then \( \gamma_n^* (q) = \gamma_n (q) \) and \( w_n^* (q) = \gamma_n (q) \) \( 1-q \leq \{ \gamma_n (q) + \nu_n \}^{1-q}. \) If \( \gamma_n (q) < \nu_n, \) then \( w_n^* (q) = \nu_n \) \( 1-q \leq \{ \gamma_n^* (q) + \nu_n \}^{1-q}. \) By (S23), (S24), and (S25), inequality (34) implies that for any constants \( 0 < \eta_1 < 1 \) and \( \eta_2 > 0, \) (25) and (30) are satisfied for sufficiently large \( n. \) The desired result follows from Corollary 5 with \( \hat{g} = g^\ast, \) because \( \lambda_{nj}^{2-q} \leq C_1^{2-q} \{ \gamma_n (q) + \nu_n \}^{2-q} \) and by (S25), \( \rho_{nj} = C_1 w_n^* (q) \{ \gamma_n^* (q) + \nu_n \} \leq C_1 \{ \gamma_n (q) + \nu_n \}^{2-q}. \)

\( \square \)

**Proof of Corollary 7.** Recall the definition \( \lambda_{nj} = C_1 \{ \gamma_n (q) + \nu_n \} \) and \( \rho_{nj} = C_1 w_n^* (q) \{ \gamma_n^* (q) + \nu_n \}. \) For a constant \( 0 < \eta_1 < 1, \) we choose and fix \( C_0' \geq C_0 \) sufficiently large, depending on \( q > 0, \) such that

\[ (\xi_1^* + 1)^2 \kappa_0^* - 2 \eta_3 \leq (C_0')^2 \eta_1^2. \]
Let \( S' = \{1 \leq j \leq p : \|g_j^s\|Q > C_0'^s\lambda_{nj}\} \). Then (25) is satisfied with \( S' \), due to (37), (S23), and the above inequality. Similarly, (30) is satisfied with \( S' \) for \( \eta_2 = \eta_3 + (C_0')^{1-q}\eta_3 \), by (S24) and simple manipulation. By Remark 8, Assumption 7 implies Assumption 5 and remains valid when \( S \) is replaced by \( S' \subset S \). The desired result follows from Corollary 5 with \( \bar{g} = g^s \).

\( \square \)

**Proof of Corollary 8.** The proof is similar to that of Corollary 6. First, we show

\[
\frac{w_n^1(q)M_F}{\gamma_n(q) + \nu_n} \leq (\nu_n)^{1-q}M_q.
\]  

(S26)

In fact, if \( \gamma_n^q(q) \geq \nu_n \), then \( \frac{w_n^1(q)}{\gamma_n(q)} = \frac{w_n^1(q)}{\gamma_n^q(q)} = \frac{w_n^1(q)}{\gamma_n^q(q)} \) and \( \frac{w_n^1(q)M_F}{\gamma_n^q(q)} = \frac{w_n^1(q)}{\gamma_n^q(q)}^{1-q}M_q \leq \{\gamma_n^q(q) + \nu_n\}^{1-q}M_q \). If \( \gamma_n(q) < \nu_n \), then \( \frac{w_n^1(q)M_F}{\gamma_n(q)} = \frac{w_n^1(q)}{\gamma_n(q)}^{1-q}M_q \leq \{\gamma_n(q) + \nu_n\}^{1-q}M_q \). By (S23), (S24), and (S26), inequality (40) implies that for any constants \( 0 < \eta_1 < 1 \) and \( \eta_2 > 0 \), (25) and (30) are satisfied for sufficiently large \( n \). The desired result follows from Corollary 5 with \( \bar{g} = g^s \), because \( \lambda_{nj}^{2-q}M_q = C_1^{2-q}\{\gamma_n(q) + \nu_n\}^{2-q}M_q \leq C_1^{2-q}\{\gamma_n(q) + \nu_n\}^{2-q}M_q \) and, by (S26), \( \rho_nM_F = C_1w_n^1(q)\{\gamma_n(q) + \nu_n\}M_F \leq C_1\{\gamma_n(q) + \nu_n\}^{2-q}M_q \).

\( \square \)

**S1.4 Proofs of Propositions 3, 4, 6, and 8**

Denote \( w_{nj} = w_{n,p+1} \) and \( \gamma_{nj} = \gamma_{n,p+1} \) for \( j = 1, \ldots, p \). By direct calculation, (60) implies that for any \( 0 \leq q \leq 1 \),

\[
\phi_n(\gamma_{n,p+1} + \nu_n)^{2-q} \leq O(1)\left\{n^{1/2}W_n\gamma_{n,p+1}^{2-q} + n^{1/2}V_n\min\left(\gamma_{n,p+1}^{1-q}, \gamma_{n,p+1}^{1-q}\nu_n\right)\right\}
\]

(S27)

where

\[
V_n = w_{n,p+1}^{\beta_n/2-\tau_0} \quad \text{and} \quad W_n = w_{n,p+1}^{\beta_n/2-\tau_0+\beta_0\tau_0/2}.
\]

We verify that the technical conditions hold as needed for Theorem 4, with \( w_{nj} = w_n^s(q) \) and \( \gamma_{nj} = \gamma_s(q) \) for \( 0 \leq q \leq 1 \). First, we verify \( \gamma_{nj} \leq w_{nj} \) for sufficiently large \( n \). It suffices to show that \( \gamma_s(q) \leq w_n^s(q) \) whenever \( \gamma_n(q) \leq 1 \) and \( \nu_n \leq 1 \). In fact, if \( \gamma_n(q) \geq \nu_n \), then \( w_n^s(q) = \gamma_n(q)^{1-q} \) and \( \gamma_s(q) = \gamma_n(q) \) \( \leq \gamma_n(q)^{1-q} \) provided \( \gamma_n(q) \leq 1 \). If \( \gamma_n(q) < \nu_n \), then \( w_n^s(q) = \nu_n^{1-q} \) and \( \gamma_s(q) = B_0^{-1/2}\nu_n^{-1-q(1-q)\beta_0/2} \leq \nu_n \leq \nu_n^{1-q} \) provided \( \nu_n \leq 1 \). Moreover, we have \( \Gamma_n\gamma_s(q)^{1-\beta_0/2} \leq 1 \) for sufficiently large \( n \), because \( \Gamma_n \) is no greater than \( O(\log^{1/2}(n)) \) and \( \gamma_n(q)^{1-\beta_0/2} \leq \gamma_n(q)^{1-\beta_0/2} \) decreases polynomially in \( n^{-1} \) for \( 0 < \beta_0 < 2 \).
Proof of Proposition 3. For $w_{nj} = 1$ and $\gamma_{nj} = \gamma_n^*(1) \asymp n^{-1/2}$, inequality (S27) with $q = 0$ and $\nu_n = o(1)$ gives

$$
\phi_n \{\gamma_n^*(1) + \nu_n\}^2 \leq O(1) \left\{ n^{1/2} \Gamma_n \gamma_n^2(1) + n^{1/2} \gamma_n^*(1) \nu_n + n \gamma_n^*(1) \nu_n^2 \right\} \\
= O(1) \left( n^{-1/2} \Gamma_n + \nu_n \right),
$$

Assumption 6(i) holds because $\Gamma_n$ is no greater than $O(\log^{1/2}(n))$. Inserting the above inequality into (29) in Corollary 4 yields the out-of-sample prediction result. The in-sample prediction result follows directly from Corollary 4.

Proof of Proposition 4. For $\gamma_{nj} = \gamma_n^*(0)$, inequality (S27) with $q = 0$ gives

$$
\phi_n \{\gamma_n^*(0) + \nu_n\}^2 \leq O(1) \left\{ n^{1/2} \Gamma_n W_n \gamma_n^2(0) + n^{1/2} V_n \gamma_n^*(0) \nu_n + n V_n^2 \gamma_n^2(0) \nu_n^2 \right\}. \tag{S29}
$$

By (S28) and $\gamma_n^*(0) = B_0^* n^{-1/2} w_n^s(0)^{q_0} / 2$, simple manipulation gives

$$
n^{1/2} V_n \gamma_n^*(0) \nu_n = B_0^* w_n^s(0)^{-1} \nu_n, \tag{S30}
$$

$$
n^{1/2} W_n \gamma_n^*(0) \nu_n^2 = B_0^* w_n^s(0)^{-1 - q_0} \gamma_n^*(0).$$

Then (43) and (S29) directly imply that Assumption 6(i) holds with any fixed $0 < \eta_0 < 1$ for sufficiently large $n$ and also (34) holds. The desired result follows from Corollary 6 with $q = 0$.

Proof of Proposition 6. For $\gamma_{nj} = \gamma_n^*(q)$, inequality (S27) with $q = 0$ gives

$$
\phi_n \{\gamma_n^*(q) + \nu_n\}^2 \leq O(1) \left\{ n^{1/2} \Gamma_n W_n \gamma_n^2(q) + n^{1/2} V_n \gamma_n^*(q) \nu_n + n V_n^2 \gamma_n^2(q) \nu_n^2 \right\}. \tag{S31}
$$

By (S28) and $\gamma_n^*(q) = B_0^* n^{-1/2} w_n^s(q)^{-q_0} / 2$, simple manipulation gives

$$
n^{1/2} V_n \gamma_n^*(q) \nu_n^{1-q} = B_0^* w_n^s(q)^{-1} \nu_n^{1-q},
$$

$$
n^{1/2} W_n \gamma_n^*(q) \nu_n^{2-2q} = B_0^* w_n^s(q)^{-1 - q_0} \gamma_n^*(q)^{1-q}.$$

Then (51) and (S31), along with the fact that $\nu_n = o(1)$, $\gamma_n(q) = o(1)$, and $q > 0$, imply that Assumption 6(i) holds with any fixed $0 < \eta_0 < 1$ for sufficiently large $n$. Moreover, (51) and (S27) with $\gamma_{nj} = \gamma_n^*(q)$ directly yield (37). The desired result follows from Corollary 7.

Proof of Proposition 8. Denote by $\gamma_{n,p+1}^+, V_n^+, W_n^+$, etc., the corresponding quantities based on $(w_{nj}', \gamma_{nj}')$. By (S28) and (S30) with $\tau_0 = 1$, we have $n^{1/2} V_n^+ \gamma_{n,p+1}^+ \nu_n = K_0^{-1} (n^{1/2} V_n \gamma_{n,p+1} \nu_n)$ and $n^{1/2} V_n \gamma_{n,p+1} \nu_n = B_0^* \min\{\nu_n \gamma_n^{-1}(0), 1\} \leq B_0^*$. Moreover, we have
\[ n^{1/2} \Gamma_n W_n^\gamma r_0^2 = n^{1/2} \Gamma_n W_n^\gamma K_0^{- \beta_0 - \gamma^2} (0) = o(1) \] for a constant \( K_0 \) independent of \( (n, p) \), because \( W_n' = 1, \Gamma_n \) is no greater than \( O(\log^{1/2}(n)) \), and \( n^{1/2} \gamma^2 (0) \) decreases polynomially in \( n^{-1} \).

For a constant \( 0 < \eta_1 < 1 \), we choose and fix \( K_0 \geq 1 \) sufficiently large, depending on \( M \) but independently of \( (n, p) \), such that Assumptions 6(i)–(ii) are satisfied, with \( (w_n, \gamma_n) \) replaced by \( (w_n', \gamma_n') \), for sufficiently large \( n \), due to (S23), (S29), and the definition \( \chi_n' = C_1 (\gamma_n' + \nu_n) \).

Moreover, by (S25), \( \rho_n' = \rho_n'(w_n') \leq K_0^{1 - \beta_0 / 2} \lambda_n w_n \leq K_0^{1 - \beta_0 / 2} C_1 (\gamma_n (0) + \nu_n)^2 \), which together with (S24) implies that (30) is satisfied for some constant \( \eta_2 > 0 \). Assumption 7 is also satisfied with \( C_0^* \) replaced by \( C_0'^* = C_0^* K_0^{\beta_0 / 2} \) and \( S \) replaced by \( \{ 1 \leq j \leq p : \| g^* \|_Q > C_0'^* \chi_n' \} \subset S \) for \( K_0 \geq 1 \) due to monotonicity in \( S \) for the validity of Assumption 7 by Remark 8, and with \( (w_n, \gamma_n) \) replaced by \( (w_n', \gamma_n') \) because (20) after the modification implies (20) itself, with \( w_n' \geq w_n \) for \( K_0 \geq 1 \) and \( \lambda_n \) constant in \( j \). The desired result follows from Corollary 5 with \( g = g^* \).

\[ \square \]

### S1.5 Proof of Proposition 5

Here we verify explicitly conditions of Theorem 5 for Sobolev spaces \( W_{r_i}^{m_i} \) and bounded variation spaces \( V_{r_i}^{m_i} \) with \( r_i = 1 \) on \([0, 1]\) in the case of \( \tau_j < 1 \), where \( \tau_i = 1 / (2m_i + 1 - 2/(r_i \wedge 2)) \). Because conditions (62), (63), (64) and (65) depend on \((m_j, r_j)\) only through \( \tau_j \), we assume without loss of generality \( 1 \leq r_j \leq 2 \). When the average marginal density of \( \{ X_i^{(j)} : i = 1, \ldots, n \} \) is uniformly bounded away from 0 and \( \infty \), the norms \( \| g_j \|_Q \) and \( \| g_j \|_{L_2} \) are equivalent, so that condition (64) and (65) hold for any \( L_2 \)-orthonormal bases \( \{ u_{\ell} : \ell \geq 1 \} \). Let \( u_0(x) \) be a mother wavelet with \( m \) vanishing moments, e.g., \( u_0(x) = 0 \) for \( |x| > c_0 \), \( \int u_0^2(x)dx = 1 \), \( \int x^m u_0(x)dx = 0 \) for \( m = 0, \ldots, \max j m_j \), and \( \{ u_{0, kl}(x) = \sqrt{2^k} u_0(2^k x - (\ell - 1)) : \ell = 1, \ldots, 2^k, k = 0, 1, \ldots \} \) is \( L_2 \)-orthonormal. We shall identify \( \{ u_{\ell} : \ell \geq 1 \} \) as \( \{ u_{0,11}, u_{0,21}, u_{0,22}, u_{0,31}, \ldots \} \). Because \( \# \{ \ell : u_{0, kl}(x) \neq 0 \} \leq 2c_0 k \) for \( x \),

\[
\sum_{\ell=2}^{2k+1-1} u_{j, \ell}^2(x) = \sum_{\ell=1}^{2k} u_{0, kl}^2(x) \leq 2c_0 2^k \| u_0 \|_\infty, \forall x,
\]

so that (62) holds. Suppose \( g_{j}(x) = \sum_{\ell=1}^{\infty} \theta_{j, \ell} u_{j, \ell}(x) = \sum_{k=0}^{\infty} \sum_{\ell=1}^{2k} \theta_{j, k} u_{0, kl}(x) \). Define \( u_{0}^{(-m)}(x) \) as the \( m \)-th integral of \( u_0 \), \( u_{0}^{(-m)}(x) = \int_{-\infty}^{x} u_0^{(-m+1)}(t) dt \), and \( g_{j}^{(m)}(x) = (d/dx)^m g_{j}(x) \). Because \( u_0 \) has vanishing moments, \( \int u_{0}^{(m)}(x) dx = 0 \) for \( m = 0, \ldots, \max j m_j \), so that \( u_{0}^{(m)}(x) = 0 \) for \( |x| > c_0 \). Due to the orthonormality of the basis functions, for \( 1 \leq \ell \leq 2^k \), we have

\[
2^{m_j k} \theta_{j, kl} = 2^{m_j k} \int g_{j}(x) u_{0, kl}(x) dx = (-1)^m \int g_{j}^{(m_j)}(x) u_{0, m_j, kl}(x) dx
\]
with \( u_{0,mk}\ell(x) = \sqrt{2^k u_0^{-m}}(2^k x - \ell + 1) \). By the Hölder inequality,

\[
\sum_{\ell=2^{k-1}}^{2^k-1} \left| 2^{m,k} \theta_{j\ell} \right|^{r_j} \leq \sum_{\ell=2^{k-1}}^{2^k-1} \int g_j^{(m_j)}(x) r_j \left| u_{0,mk}\ell(x) \right|^{r_j(1-2(1-r_j))} dx \left| u_{0,mk}\ell \right|^{r_j(2-1(1-r_j))} \left| L_2 \right|
\]

\[
\leq \left| g_j^{(m_j)} \right|^{r_j} \left| L_{r_j} \right| 2c_0 \left| u_0^{(m_j)} \right|^{r_j(1-2(1-r_j))} 2^{(1-2(1-r_j))} \left| u_0^{(m_j)} \right|^{2(1-1(1-r_j))} \left| L_2 \right|.
\]

Because \( 2^{m_j,k-2(1-r_j)} = 2^{k(m_j+1-2(1-r_j))} = 2^{(2-r_j)} \) and \( 1 \leq r_j \leq 2 \), we have

\[
\left\{ \sum_{\ell=2^{k-1}}^{2^k-1} 2^{k/r_j} \theta_{j\ell}^2 \right\}^{1/2} \leq \left\{ \sum_{\ell=2^{k-1}}^{2^k-1} 2^{(2/r_j)} \theta_{j\ell}^2 \right\}^{1/2} \leq \left| g_j^{(m_j)} \right|^{1/r_j} \left| L_{r_j} \right| (2c_0)^{1/r_j} \left| u_0^{(m_j)} \right|^{2/r_j-1} \left| u_0^{(m_j)} \right|^{2-2/r_j} \left| L_2 \right|.
\]

Because \( \ell_{j/k}^{1/(2r_j)} \geq 2^{k/w_{nj}} \) and \( \ell_{jk} \leq 1 + 2^{2r_j} \ell_{j,k-1} \) with \( r_j < 1 \), we have \( \ell_{jk} \leq 4\ell_{j,k-1} \), so that \( \{\ell_{j,k-1} + 1, \ldots, \ell_{j,k}\} \) involves at most three resolution levels. Thus, condition (63) follows from the above inequality. For the bounded variation class, we have

\[
2^{m_j,k} \theta_{j\ell} = 2^{m_j,k} \int g_j(x) u_{0,k\ell}(x) dx = (-1)^m \int u_{0,mk\ell}(x) dg_j^{(m_j-1)}(x),
\]

so that (63) follows from the same proof with \( r_j = 1 \).

### S1.6 Proof of Theorem 3

(i) For \( \delta_j \in \{0, 1\} \) and \( e_{j\ell} \in \{-1, 1\} \), let

\[
g_j(t) = \sigma_n \sum_{\ell=1}^{k} \delta_j e_{j\ell} u_{j\ell}(t), \quad j = 1, \ldots, p,
\]

where \( \sigma_n = \sigma_n^{-1/2} \). By (53) and (54),

\[
\|g_j\|_{Q}^2 \leq \sum_{\ell=1}^{k} (\delta_j \sigma_n e_{j\ell})^2 = \delta_j \sigma_n^2 k, \quad \|g_j\|_{F,j} \leq \delta_j \sigma_n C_F k^{1/3n+1/2}.
\]

(S32)

Consider probability \( \mathbb{P} \) under which \( \{\delta_j : 1 \leq j \leq p\} \) are deterministic and \( \{e_{j\ell} : 1 \leq \ell \leq k, 1 \leq j \leq p\} \) are independent variables with

\[
\delta_j \in \{0, 1\}, \quad \sum_{j=1}^{p} \delta_j = s, \quad \mathbb{P}\{e_{j\ell} = \pm 1\} = 1/2.
\]

(S33)

It follows from (55), (S32) and (S33) that

\[
\sum_{j=1}^{p} \|g_j\|_{Q}^2 \leq s \sigma_n^2 k^{3/2} = \sigma^2 M_q, \quad \sum_{j=1}^{p} \|g_j\|_{F,j} \leq s \sigma_n C_F k^{1/3n+1/2} = \sigma M_F,
\]

13
so that $\mathbb{P}\{g \in \mathcal{G}(M_F, M_g)\} = 1$.

It suffices to find a lower bound for the Bayes risk of the Bayes estimator. Denote $X = (X_1, \ldots, X_n)^\top = (X_{i1}^{(j)})_{n \times p}$ and $Y = (Y_1, \ldots, Y_n)^\top$. For the estimation of $g_j$, the risk of the Bayes rule is bounded by

$$
\mathbb{E}\|g_j - \mathbb{E}(g_j | X, Y)\|^2_Q \geq c_0 \mathbb{E} \sum_{\ell=1}^k \sigma^2_{\ell} \text{Var}\left(\delta_{j, \ell} | X, Y\right).
$$

(S34)

Let $Y_{j\ell} = (Y_{ij} : 1 \leq i \leq n)^\top$ with $Y_{ij} = \sigma_n \delta_{j, \ell} u_{j\ell}(X_{i1}^{(j)}) + \epsilon_i$. We have

$$
\mathbb{E}\text{Var}\left(\delta_{j, \epsilon_{j\ell}} | X, Y\right) \geq \mathbb{E}\text{Var}\left(\delta_{j, \epsilon_{j\ell}} | X, Y, \epsilon_{j', \ell'}, (j', \ell') \neq (j, \ell)\right)
= \mathbb{E}\text{Var}\left(\delta_{j, \epsilon_{j\ell}} | X, Y_{j\ell}\right)
= \delta_j \mathbb{E}\left\{1 - \mathbb{E}^2 \left(\epsilon_{j\ell} | X, Y_{j\ell}\right)\right\}.
$$

(S35)

Let $v_{j\ell} = \sum_{i=1}^n u_{i1}^2(X_{i1}^{(j)})$, $\mu_{j\ell} = \sqrt{v_{j\ell}/n}$ and $z_{j\ell} = \sum_{i=1}^n u_{ij}(X_{i1}^{(j)})Y_{ijk}/v_{j\ell}^{1/2}$. Given $\delta_j = 1$ and $(X, Y_{j\ell})$, $z_{j\ell}$ is sufficient for $\epsilon_{j\ell}$. As $\delta_j z_{j\ell}(\epsilon_{j\ell}, \mu_{j\ell}) \sim \delta_j N(\epsilon_{j\ell} \sigma_{\mu_{j\ell}}, \sigma^2)$, the posterior of $\{\epsilon_{j\ell} : 1 \leq \ell \leq m\}$ given $\delta_j = 1$ and $(X, Y_{j\ell})$ is proportional to

$$
\exp\left\{-\delta_j \frac{\left(\sum_{\ell=1}^k \frac{(z_{j\ell} - \epsilon_{j\ell}|\mu_{j\ell}|)^2}{2\sigma^2}\right)}{\sigma^2}\right\} \propto \exp\left\{\sum_{\ell=1}^k \frac{z_{j\ell}(\epsilon_{j\ell}|\mu_{j\ell}|)}{\sigma^2}\right\}.
$$

It follows that

$$
\delta_j \mathbb{E}\left(\epsilon_{j\ell} | X, Y_{j\ell}\right) = \delta_j \frac{\sinh(z_{j\ell}(\mu_{j\ell}/\sigma))}{\cosh(z_{j\ell}(\mu_{j\ell}/\sigma))}.
$$

For $\delta_j = 1$ and given $(\epsilon_{j\ell}, \mu_{j\ell})$, $z_{j\ell}/\sigma \sim N(\epsilon_{j\ell}\mu_{j\ell}, 1)$ and hence

$$
\delta_j \mathbb{E}\left\{1 - \mathbb{E}^2 \left(\epsilon_{j\ell} | X, Y_{j\ell}\right) | \mu_{j\ell}\right\}
= \delta_j \int \left(1 - \frac{\sinh^2(\mu_{j\ell}x)}{\cosh^2(\mu_{j\ell}x)}\right) \cosh(\mu_{j\ell}x) e^{-\mu_{j\ell}^2/2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx
= \delta_j \int \frac{e^{-\mu_{j\ell}^2/2}}{\cosh(\mu_{j\ell}x)} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx
\geq \delta_j \int \frac{1}{\sqrt{2\pi}} e^{-\mu_{j\ell}^2/2 - \mu_{j\ell}|x|} e^{-x^2/2} dx = \delta_j c_1.
$$

Consequently, as $\mathbb{E} \mu_{j\ell}^2 = \mathbb{E} v_{j\ell}/n = \|u_{j\ell}\|^2_Q \leq 1$,

$$
\delta_j \mathbb{E}\left\{1 - \mathbb{E}^2 \left(\epsilon_{j\ell} | X, Y_{j\ell}\right)\right\} \geq \delta_j \int \frac{1}{\sqrt{2\pi}} e^{-1/2 - |x|} e^{-x^2/2} dx = \delta_j c_1.
$$

It follows from (S34), (S35) and the above inequality that

$$
\mathbb{E} \sum_{j=1}^p \left\|g_j - \mathbb{E}(g_j | X, Y)\right\|^2_Q \geq c_0 \sum_{j=1}^p \delta_j c_1 k \sigma_n^2 \geq c_0 c_1 s k \sigma_n^2.
$$

14
It remains to compute $sk\sigma_n^2$. By (55), we have

$$s = \left( \frac{n^{1/2} M_F}{C_F} \right)^{-q\beta_0/(2+(1-q)\beta_0)} \left( \frac{n^{q/2} M_q}{M_F} \right)^{(2+\beta_0)/(2+(1-q)\beta_0)},$$

$$k = \left( \frac{n^{1/2} M_F}{C_F} \right)^{2\beta_0/(2+(1-q)\beta_0)} \left( \frac{n^{q/2} M_q}{M_F} \right)^{-2\beta_0/(2+(1-q)\beta_0)},$$

and hence $sk\sigma_n^2 = \sigma^2(M_F/C_F)^{1-\eta} M_q^{1+\eta}(2-q)/(2+(1-q)\beta_0)$.

(ii) Let $N$ be a uniform random variable in $\{0, \ldots, s_0\}$, and given $N$ let $\delta_j \in \{0, 1\}$ such that $\{j : \delta_j = 1\}$ is a simple random sample (without replacement) of size $N$ in $\{1, \ldots, p\}$. Let $\tau = 1/2$ and

$$g_j(t) = \sigma \lambda_0 \delta_j \sqrt{1-\tau} u_j(t), \quad j = 1, \ldots, p.$$ 

As $\sum_j \delta_j = N \leq s_0 \leq \min (M_q/\lambda_0^2, M_F/(C_F\lambda_0), p)$,

$$\sum_j \|g_j\|_q^q \leq \sigma^q N \lambda_0^q \leq \sigma^q M_q, \quad \sum_j \|g_j\|_{F,j} \leq \sigma N C_F \lambda_0 \leq \sigma M_F.$$

Similar to the proof of part (i), the Bayes risk of the Bayes rule is bounded from below by

$$\mathbb{E}\left[ \left\| g_j - \mathbb{E}(g_j|X, Y) \right\|_Q^2 \right] \geq \mathbb{E} \var\left( \delta_j | X, Y_{j1}, N_j \right) (1-\tau) \sigma^2 \lambda_0^2$$

$$= \mathbb{E} \var\left( \delta_j | z_{j1}, v_{j1}, N_j \right) (1-\tau) \sigma^2 \lambda_0^2,$$

where $Y_{j1}, z_{j1}$ and $v_{j1}$ are as in the proof of (i), and $N_j = N - \delta_j$. The difference here is $k = e_{j1} = 1, n_0 = 2 \log(ep/s_0)$ and $\{\delta_j : 1 \leq j \leq p\}$ are random. We still have

$$z_{j1}(\delta_j, v_{j1}, N_j) \sim N(\delta_j v_{j1}^{1/2} \sqrt{1-\tau} \sigma \lambda_0, \sigma^2).$$

Moreover, for $0 \leq k < s_0$, we have

$$P(N_j = k) = \frac{(1-k/p) + (k+1)/p}{s_0 + 1} = \frac{1 + 1/p}{s_0 + 1},$$

$$\pi_k = P(\delta_j = 1 | N_j = k) = \frac{(k+1)/p}{1 + 1/p} = \frac{k + 1}{p + 1}.$$

Let $\pi_0 = s_0/p$ and $\mu_{j1} = v_{j1}^{1/2} \sqrt{1-\tau} \sigma \lambda_0$. For $0 \leq k < s_0$ and $(2z_{\mu_{j1}} - \mu_{j1}^2)/(2\sigma^2) \leq n \lambda_0^2$, we have $\pi_k \leq \pi_0$ and

$$\mathbb{E}\left( \delta_j | z_{j1} = z, v_{j1}, N_j = k \right) = \pi_k \exp\left\{ -z \mu_{j1}/(2\sigma^2) \right\}$$

$$= \frac{1 - \pi_k}{\pi_0} \exp\left\{ -z \mu_{j1}/(2\sigma^2) \right\} + \pi_k \exp\left\{ -z \mu_{j1}/(2\sigma^2) \right\}$$

$$\leq \frac{1}{1 - \pi_0} + \pi_0 \exp\left\{ (2z_{\mu_{j1}} - \mu_{j1}^2)/(2\sigma^2) \right\}$$

$$\leq \frac{1}{1 - s_0/p + 1/e}.$$

15
By direct calculation, it follows that

$$E \text{Var}(\delta_j | z_{j1}, v_{j1}, N_j)$$

$$= E \left\{ \left( 1 - E(\delta_j | z_{j1}, v_{j1}, N_j) \right) E(\delta_j | z_{j1}, v_{j1}, N_j) \right\}$$

$$\geq \frac{1 - s_0/p}{1 - s_0/p + 1/e} E \left\{ E(\delta_j | z_{j1}, v_{j1}, N_j) I\{ (2z_{j1}\mu_{j1} - \mu_{j1}^2)/\sigma^2 \leq n\lambda_0^2, N_j < s_0 \} \right\}$$

$$= \frac{1 - s_0/p}{1 - s_0/p + 1/e} \sum_{k=0}^{s_0-1} P(N_j = k) P(\delta_j = 1 | N_j = k)$$

$$\times P\left\{ (2z_{j1}\mu_{j1} - \mu_{j1}^2)/\sigma^2 \leq n\lambda_0^2 | \delta_j = 1, N_j = k \right\}.$$

As $2z_{j1}\mu_{j1} - \mu_{j1}^2$ given $\mu_{j1}$ and $\delta_j = 1$ is Gaussian with mean $\mu_{j1}^2$, the above inequality implies

$$E \text{Var}(\delta_j | z_{j1}, v_{j1}, N_j)$$

$$\geq \frac{1 - s_0/p}{1 - s_0/p + 1/e} \sum_{k=0}^{s_0-1} \left( 1 + 1/p \right) \left( \frac{k + 1}{s_0 + 1} \right) \frac{1}{2} P(\mu_{j1}^2/\sigma^2 \leq n\lambda_0^2)$$

$$= (1 - s_0/p)s_0/(4p) \sum_{i=1}^{n} u_{j1}^2(X_{ij})(1 - \tau) \leq n \right\}$$

$$\geq (1 - s_0/p)s_0/(4p) \left\{ 1 - (1 - \tau)E u_{j1}^2(X_{ij}) \right\}.$$

Consequently, the Bayes risk of the Bayes rule is bounded from below by

$$\sum_{j=1}^{p} \left\| g_j - E[g_j | X, Y] \right\|^2_Q \geq \frac{(1 - s_0/p)\tau/4}{1 - s_0/p + 1/e} s_0 (1 - \tau)\sigma^2\lambda_0^2.$$

The conclusion follows as we have picked $\tau = 1/2$.

**S1.7 Proof of Theorem 4**

We split the proof into three lemmas. First, we provide maximal inequalities on convergence of empirical inner products in functional classes with polynomial entropies.

**Lemma 6.** Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two functional classes such that

$$\sup_{f_j \in \mathcal{F}_j} \|f_j\|_Q \leq \delta_j, \quad \sup_{f_j \in \mathcal{F}_j} \|f_j\|_\infty \leq b_j, \quad j = 1, 2.$$

Suppose that for some $0 < \beta_j < 2$ and $B_{n_j,\infty} > 0$, condition (S49) holds with

$$\psi_{n,\infty}(z, \mathcal{F}_j) = B_{n_j,\infty} z^{1-\beta_j/2}, \quad j = 1, 2.$$  

(S36)
Then we have
\[
E \left\{ \sup_{f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2} |\langle f_1, f_2 \rangle_n - \langle f_1, f_2 \rangle_q| / C_2 \right\} \\
\leq 2 \left\{ \delta_1 + \frac{2C_2 \psi_{n,\infty}(b_1, \mathcal{F}_1)}{\sqrt{n}} \right\}^{1-\beta_1/2} \left\{ \delta_2 + \frac{2C_2 \psi_{n,\infty}(b_2, \mathcal{F}_2)}{\sqrt{n}} \right\}^{\beta_1/2} \psi_{n,\infty}(b_1, \mathcal{F}_1) \\
+ 2 \left\{ \delta_2 + \frac{2C_2 \psi_{n,\infty}(b_2, \mathcal{F}_2)}{\sqrt{n}} \right\}^{1-\beta_2/2} \left\{ \delta_1 + \frac{2C_2 \psi_{n,\infty}(b_1, \mathcal{F}_1)}{\sqrt{n}} \right\}^{\beta_2/2} \psi_{n,\infty}(b_1, \mathcal{F}_2). \tag{S37}
\]
Moreover, we have for any \( t > 0 \),
\[
\sup_{f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2} |\langle f_1, f_2 \rangle_n - \langle f_1, f_2 \rangle_q| / C_3 \\
\leq E \left\{ \sup_{f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2} |\langle f_1, f_2 \rangle_n - \langle f_1, f_2 \rangle_q| \right\} + \delta_1 b_2 \sqrt{\frac{t}{n} + b_1 b_2 \frac{t}{n}}, \tag{S38}
\]
with probability at least \( 1 - e^{-t} \).

**Proof.** For any function \( f_1, f'_1 \in \mathcal{F}_1 \) and \( f_2, f'_2 \in \mathcal{F}_2 \), we have by triangle inequalities,
\[
||f_1 f_2 - f'_1 f'_2||_n \leq \delta_2 ||f_1 - f'_1||_{n,\infty} + \delta_1 ||f_2 - f'_2||_{n,\infty}.
\]
As a result, we have for \( u > 0 \),
\[
H(u, \mathcal{F}_1 \times \mathcal{F}_2, \| \cdot \|_n) \leq H\{u/(2\delta_2), \mathcal{F}_1, \| \cdot \|_{n,\infty}\} + H\{u/(2\delta_1), \mathcal{F}_2, \| \cdot \|_{n,\infty}\}, \tag{S39}
\]
where \( \mathcal{F}_1 \times \mathcal{F}_2 = \{f_1 f_2 : f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2\} \).

By symmetrization inequality (van der Vaart & Wellner 1996),
\[
E \left\{ \sup_{f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2} |\langle f_1, f_2 \rangle_n - \langle f_1, f_2 \rangle_q| \right\} \leq 2E \left\{ \sup_{f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2} |\langle \sigma, f_1 f_2 \rangle_n| \right\}.
\]
Let \( \hat{\delta}_1 = \sup_{f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2} ||f_1 f_2||_n \leq \min(\delta_1 b_2, \delta_2 b_1) \). By Dudley’s inequality (Lemma 13) conditionally on \( X_{1:n} = (X_1, \ldots, X_n) \), we have
\[
E \left\{ \sup_{f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2} |\langle \sigma, f_1 f_2 \rangle_n| X_{1:n} \right\} / C_2 \leq E \left\{ \int_0^{\hat{\delta}_1} H^{1/2}(u, \mathcal{F}_1 \times \mathcal{F}_2, \| \cdot \|_n) \, du \right\}.
\]
Taking expectations over \( X_{1:n} \), we have by (S39), (S49), and definition of \( H^*(\cdot) \),
\[
E \left\{ \sup_{f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2} |\langle f_1, f_2 \rangle_n - \langle f_1, f_2 \rangle_q| \right\} / C_2 \\
\leq E \left[ \int_0^{\hat{\delta}_1 b_2} H^{1/2}\{u/(2\delta_2), \mathcal{F}_1, \| \cdot \|_{n,\infty}\} \, du + \int_0^{\hat{\delta}_1 b_1} H^{1/2}\{u/(2\delta_1), \mathcal{F}_2, \| \cdot \|_{n,\infty}\} \, du \right] \\
\leq E \left[ \hat{\delta}_2 \psi_{n,\infty}(\hat{\delta}_1 b_2/\hat{\delta}_2, \mathcal{F}_1) + \hat{\delta}_1 \psi_{n,\infty}(\hat{\delta}_2 b_1/\hat{\delta}_1, \mathcal{F}_2) \right]. \tag{S40}
\]
By (S36) and the Hölder inequality, we have
\[
E \left\{ \tilde{\delta}_2 \psi_{n,\infty}(\tilde{\delta}_1 b_1/\tilde{\delta}_2, F_1) \right\} \leq B_{n1,\infty} b_2^{1-\beta_1/2} E \left( \tilde{\delta}_2^{\beta_1/2} \tilde{\delta}_1^{1-\beta_1/2} \right)
\]
\[
\leq B_{n1,\infty} b_2^{1-\beta_1/2} E^{\beta_1/2}(\tilde{\delta}_2) E^{1-\beta_1/2}(\tilde{\delta}_1) \leq B_{n1,\infty} b_2^{1-\beta_1/2} E^{\beta_1/4}(\tilde{\delta}_2) E^{(2-\beta_1)/4}(\tilde{\delta}_1),
\]
and similarly
\[
E \left\{ \tilde{\delta}_1 \psi_{n,\infty}(\tilde{\delta}_2 b_1/\tilde{\delta}_1, F_2) \right\} \leq B_{n2,\infty} b_1^{1-\beta_2/2} E^{\beta_2/4}(\tilde{\delta}_1) E^{(2-\beta_2)/4}(\tilde{\delta}_2).
\]
Then inequality (S37) follows from (S40) and Lemma 16. Moreover, inequality (S38) follows from Talagrand’s inequality (Lemma 14) because \(\|f_1 f_2\|_Q \leq \delta_1 b_2\) and \(\|f_1 f_2\|_\infty \leq b_1 b_2\) for \(f_1 \in F_1\) and \(f_2 \in F_2\).

By application of Lemma 6, we obtain the following result on uniform convergence of empirical inner products under conditions (57), (58), and (59).

**Lemma 7.** Suppose the conditions of Theorem 4 are satisfied for \(j = 1, 2\) and \(p = 2\). Let \(F_j = G_j^*(w_{nj})\) for \(j = 1, 2\). Then we have
\[
E \left\{ \sup_{f_1 \in F_1, f_2 \in F_2} |\langle f_1, f_2 \rangle_n - \langle f_1, f_2 \rangle_Q| / C_2 \right\} \leq 2(1 + 2C_2C_4)C_4 n^{1/2} \Gamma_n \sigma_{w_{nj}}^2 \left( \gamma_{n1} \tilde{\gamma}_{n2} w_{nj1}^{\beta_1 \tau_2} + \gamma_{n2} \tilde{\gamma}_{n1} w_{nj1}^{\beta_2 \tau_1} \right),
\]
where \(0 < \tau_j \leq (2/\beta_j - 1)^{-1}\) and \(C_4 = \max_{j=1,2} C_{4,j}\) from condition (59), and \(\tilde{\gamma}_{nj} = n^{-1/2} w_{nj}^{-\tau_j}\). Moreover, we have for any \(t > 0\),
\[
\sup_{f_1 \in F_1, f_2 \in F_2} |\langle f_1, f_2 \rangle_n - \langle f_1, f_2 \rangle_Q| / C_3 \leq E \left\{ \sup_{f_1 \in F_1, f_2 \in F_2} |\langle f_1, f_2 \rangle_n - \langle f_1, f_2 \rangle_Q| \right\} + w_{nj1} \left( C_4 t^{1/2} \tilde{\gamma}_{nj} + C_4^2 t^{\beta_1 \tau_2} \right),
\]
with probability at least 1 - \(e^{-t}\).

**Proof.** For \(f_j \in F_j\) with \(w_{nj} \leq 1\), we have \(\|f_j\|_{F_j} \leq 1\) and \(\|f_j\|_Q \leq w_{nj}\), and hence \(\|f_j\|_\infty \leq C_4 w_{nj}^{1-\tau_j}\) by (59). Let \(\psi_{n,\infty}(\cdot, F_j) = \psi_{n,\infty}(\cdot, w_{nj})\) from (58), that is, in the form (S36) such that (S49) is satisfied. We apply Lemma 6 with \(\delta_j = w_{nj}\) and \(b_j = C_4 w_{nj}^{1-\tau_j}\). By simple manipulation, we have
\[
n^{-1/2} \psi_{n,\infty}(b_j, F_j) = n^{-1/2} \psi_{nj,\infty}(C_4 w_{nj}^{1-\tau_j}, w_{nj})
\]
\[
\leq C_4 B_{nj,\infty} n^{-1/2} w_{nj}^{-\beta_j/2} w_{nj}^{1-(1-\beta_j/2)\tau_j} \leq C_4 \Gamma_n \gamma_{nj} w_{nj}^{1-\beta_j/2} \leq C_4 w_{nj},
\]

18
where $C_4 \geq 1$ is used in the second step, $B_{n_j, \infty} \leq \Gamma_n B_{n_j}$ and $(1 - \beta_j/2) \tau_j \leq \beta_j/2$ in the third step, and $\gamma_{n_j} \leq w_{n_j}$ and $\Gamma_n \gamma_{n_j} w_{n_j}^{-\beta_j/2} \leq \Gamma_n \gamma_{n_j}^{1-\beta_j/2}$ $\leq 1$ in the fourth step. Therefore, inequality (S37) yields

$$E \left\{ \sup_{f_1 \in F_1, f_2 \in F_2} |(f_1, f_2)_n - (f_1, f_2)_Q| / C_2 \right\}$$

$$\leq 2(1 + 2C_2 C_4) n^{-1/2} w_{n_1}^{1-\beta_1/2} w_{n_2}^{\beta_2/2} \psi_{n_1, \infty}(C_4 w_{n_2}^{1-\tau_2}, w_{n_1})$$

$$+ 2(1 + 2C_2 C_4) n^{-1/2} w_{n_2}^{1-\beta_2/2} w_{n_1}^{\beta_2/2} \psi_{n_2, \infty}(C_4 w_{n_1}^{1-\tau_1}, w_{n_2})$$

$$\leq 2(1 + 2C_2 C_4) C_4 n^{-1/2} w_{n_1}^{1-\beta_1/2} B_{n_1, \infty} w_{n_2}^{\tau_2} + \beta_2 \tau_2$$

$$+ 2(1 + 2C_2 C_4) C_4 n^{-1/2} w_{n_2}^{1-\beta_2/2} B_{n_2, \infty} w_{n_1}^{\tau_1} + \beta_1 \tau_1 / 2,$$

which leads to the first desired inequality because $B_{n_j, \infty} \leq \Gamma_n B_{n_j}$. Moreover, simple manipulation gives

$$\delta_1 b_2 \sqrt{\frac{t}{n}} = C_4 w_{n_1} w_{n_2}^{1-\tau_2} \sqrt{\frac{t}{n}} = C_4 t^{1/2} w_{n_1} w_{n_2} \tilde{\gamma}_{n_2},$$

$$b_1 b_2 \frac{t}{n} = C_4^2 w_{n_1} \tilde{\gamma}_{n_1} w_{n_2} \frac{t}{n} = C_4^2 t w_{n_1} \tilde{\gamma}_{n_1} w_{n_2} \tilde{\gamma}_{n_2}.$$

The second desired inequality follows from (S38).

The following result concludes the proof of Theorem 4.

**Lemma 8.** In the setting of Theorem 4, let

$$\phi_n = 4C_2 C_3 (1 + 2C_2 C_4) C_4 n^{1/2} \Gamma_n \max_j \frac{\gamma_{n_j}}{\lambda_{n_j}} \max_j \frac{\tilde{\gamma}_{n_j} w_{n_j}^{\beta_j+\tau_j/2}}{\lambda_{n_j}^2}$$

$$+ \sqrt{2} C_3 C_4 \max_j \frac{\tilde{\gamma}_{n_j}}{\lambda_{n_j}} \max_j \sqrt{\frac{\log(p/\epsilon')}{\lambda_{n_j}}} + 2C_3 C_4^2 \max_j \frac{\tilde{\gamma}_{n_j}^2 \log(p/\epsilon')}{\lambda_{n_j}^2},$$

where $\tilde{\gamma}_{n_j} = n^{-1/2} w_{n_j}^{-\tau_j}$ and $\beta_{p+1} = \min_{j=1, \ldots, p} \beta_j$. Then

$$P \left\{ \sup_{g \in \mathcal{G}} \frac{\|g\|_n^2 - \|g\|_Q^2}{R_n^2(g)} > \phi_n \right\} \leq \epsilon'^2.$$

**Proof.** For $j = 1, \ldots, p$, let $r_{n_j}^*(g_j) = \|g_j\|_{F,j} + \|g_j\|_Q / w_{n_j}$ and $f_j = g_j / r_{n_j}^*(g_j)$. Then $\|f_j\|_{F,j} + \|f_j\|_Q / w_{n_j} = 1$ and hence $f_j \in \mathcal{G}_n^*(w_{n_j})$. By the decomposition $\|g\|_n^2 = \sum_{j,k} \langle g_j, g_k \rangle_n$, $\|g\|_Q^2 = \sum_{j,k} \langle g_j, g_k \rangle_Q$, and the triangle inequality, we have

$$\|g\|_n^2 - \|g\|_Q^2 \leq \sum_{j,k} |\langle g_j, g_k \rangle_n - \langle g_j, g_k \rangle_Q|$$

$$= \sum_{j,k} r_{n_j}^*(g_j) r_{n_k}^*(g_k) |\langle f_j, f_k \rangle_n - \langle f_j, f_k \rangle_Q|.$$
Because $R^{*2}(g) = \sum_{j,k} r_{nj}^*(g_j)r_{nk}^*(g_k)w_{nj}\lambda_{nj}w_{nk}\lambda_{nk}$, we have

$$\left\{ \sup_{g = \sum_{j=1}^p g_j} \frac{\|g\|_2^2 - \|g\|_Q^2}{R^{*2}(g)} > \phi_n \right\} = \bigcup_{g = \sum_{j=1}^p g_j} \left\{ \|g\|_2^2 - \|g\|_Q^2 > \phi_n R^{*2}(g) \right\}$$

$$\subset \bigcup_{j,k} \left\{ \sup_{f_j \in G^*(w_{nj}), f_k \in G^*(w_{nk})} |\langle f_j, f_k \rangle_n - \langle f_j, f_k \rangle_Q| > \phi_n w_{nj}\lambda_{nj}w_{nk}\lambda_{nk} \right\}$$

By Lemma 7 with $F_1 = G_{\epsilon}^*(w_{nj})$, $F_2 = G_{\epsilon}^*(w_{nk})$, and $t = \log(p^2/\epsilon^2)$, we have with probability no greater than $\epsilon^2/p^2$,

$$\sup_{f_j \in G^*(w_{nj}), f_k \in G^*(w_{nk})} |\langle f_j, f_k \rangle_n - \langle f_j, f_k \rangle_Q|/C_3$$

$$> 4C_2(1 + 2C_2C_4)C_4n^{1/2}\Gamma_nw_{nj}\gamma_{nj}w_{nk}\gamma_{nk}$$

$$+ C_4n^{1/2}V_nw_{nj}\gamma_{nj}w_{nk}\gamma_{nk}\sqrt{\log(p^2/\epsilon^2)/n} + C_4^3V_n^2\log(p^2/\epsilon^2)w_{nj}\gamma_{nj}w_{nk}\gamma_{nk}.$$ 

Therefore, we have by the definition of $\phi_n$,

$$P\left( \sup_{f_j \in G^*(w_{nj}), f_k \in G^*(w_{nk})} |\langle f_j, f_k \rangle_n - \langle f_j, f_k \rangle_Q| > \phi_n w_{nj}\lambda_{nj}w_{nk}\lambda_{nk} \right) \leq \frac{\epsilon^2}{p^2}.$$ 

The desired result follows from the union bound. \qed

### S1.8 Proof of Theorem 5

We use the non-commutative Bernstein inequality (Lemma 15) to prove Theorem 5. Suppose that $(X_1, \ldots, X_n)$ are independent variables in a set $\Omega$. First, consider finite-dimensional functional classes $F_j$ with elements of the form

$$f_j(x) = u_j^T(x)\theta_j, \quad \forall \theta_j \in \mathbb{R}^{d_j}, j = 1, 2,$$

where $u_j(x)$ is a vector of basis functions from $\Omega$ to $\mathbb{R}^{d_j}$, and $\theta_j$ is a coefficient vector. Let $U_j = \{u_j(X_1), \ldots, u_j(X_n)\}^\top$, and $\Sigma_{jj'} = E(U_j^\top U_{j'}/n) \in \mathbb{R}^{d_j \times d_{j'}}$. The population inner product is $\langle f_j, f_j' \rangle_Q = \theta_j^\top\Sigma_{jj'}\theta_{j'}$, $j, j' = 1, 2$. The difference between the sample and population inner products can be written as

$$\sup_{\|\theta_j\| = \|\theta_{j'}\| = 1} |\langle f_j, f_j' \rangle_n - \langle f_j, f_j' \rangle_Q| = \sup_{\|\theta_j\| = \|\theta_{j'}\| = 1} |\theta_j^\top(U_j^\top U_{j'}/n - \Sigma_{jj'})\theta_{j'}|$$

$$= \|U_j^\top U_{j'}/n - \Sigma_{jj'}\|_{S}.$$
Lemma 9. Let \( f_j \) be as in (S41). Assume that for a constant \( C_{5,1} \),

\[
\sup_{x \in \Omega} \| u_j(x) \|^2 \leq C_{5,1} \ell_j, \quad \forall j = 1, 2.
\]

Then for all \( t > 0 \),

\[
\left\| U_j^T U_j' / n - \Sigma_{jj'} \right\| S > \sqrt{\left( \ell_j \| \Sigma_{jj'} \| S \right) \vee \left( \ell_j' \| \Sigma_{jj'} \| S \right)} \sqrt{\frac{2C_{5,1} t}{n}} + C_{5,1} \sqrt{\ell_j \ell_j'} \frac{4t}{3n}
\]

with probability at least \( 1 - (d_j + d_j') e^{-t} \).

Proof. Let \( M_i = u_j(X_i) u_j'(X_i) - E\{ u_j(X_i) u_j'(X_i) \} \). Because \( u_j(X_i) u_j'(X_i) \) is of rank 1, \( \| M_i \| S \leq 2 \sup_{x \in \Omega} \{ \| u_j(x) \| \| u_j'(x) \| \} \leq 2C_{5,1} \sqrt{\ell_j \ell_j'} \). Hence we set \( s_0 = 2C_{5,1} \sqrt{\ell_j \ell_j'} \) in Lemma 15. Similarly, \( W_{\text{col}} \leq C_{5,1} \ell_j \| \Sigma_{jj} \| S \) because

\[
E(M_i M_i^T) \leq E\{ u_j(X_i) u_j'(X_i) u_j(X_i) u_j'(X_i) \} \leq C_{5,1} \ell_j E\{ u_j(X_i) u_j'(X_i) \},
\]

and \( W_{\text{row}} \leq C_{5,1} \ell_j \| \Sigma_{jj'} \| S \). Thus, (S48) gives the desired result. \( \square \)

Now consider functional classes \( \mathcal{F}_j \) such that \( f_j \in \mathcal{F}_j \) admits an expansion

\[
f_j(\cdot) = \sum_{\ell=1}^{\infty} \theta_{j,\ell} u_{j,\ell}(\cdot),
\]

where \( \{ u_{j,\ell}(\cdot) : \ell = 1, 2, \ldots \} \) are basis functions and \( \{ \theta_{j,\ell} : \ell = 1, 2, \ldots \} \) are the associated coefficients.

Lemma 10. Let \( 0 < \tau_j < 1 \), \( 0 < w_{nj} \leq 1 \) and

\[
B_j = \left\{ f_j : \sum_{k/4 < \ell \leq k} \theta_{j,\ell}^2 \leq k^{-1/\tau_j} \forall k \geq \left( 1 / w_{nj} \right)^{2\tau_j}, \sum_{0 \leq \ell \leq \tau_j} \theta_{j,\ell+1}^2 \leq w_{nj}^2 \right\}
\]

Suppose that (62) and (65) hold with certain positive constants \( C_{5,1}, C_{5,3} \). Then, for a certain constant \( C_{5,4} \) depending on \( \{ C_{5,1}, C_{5,3} \} \) only,

\[
\sup_{f_j \in B_j, f_j' \in B_j'} \left| \langle f_j, f_j' \rangle_n - \langle f_j, f_j' \rangle_Q \right|
\]

\[
\leq C_{5,4} w_{nj} w_{nj'} \left( \mu_j w_{nj}^{-\tau_j} + \mu_j' w_{nj'}^{-\tau_j'} \right) \sqrt{\left\{ \mu_j + \mu_j' + \log( w_{nj}^{-\tau_j} + w_{nj'}^{-\tau_j'}) + t \right\} / n}
\]

\[
+ \left\{ \mu_j + \mu_j' + \log( w_{nj}^{-\tau_j} + w_{nj'}^{-\tau_j'}) + t \right\} \left( \mu_j w_{nj}^{-\tau_j} (\mu_j' w_{nj'}^{-\tau_j'}) / n \right)
\]

with at least probability \( 1 - e^{-t} \) for all \( t > 0 \), where \( \mu_j = 1 / (1 - \tau_j) \).
Proof. Let $\ell_{jk} = [(2^k/w_{nj})^{2\tau_j}]$. We group the basis and coefficients as follows:

$$u_{j,G_{jk}}(x) = (u_{j\ell}(x), \ell \in G_{jk})^T, \quad \theta_{j,G_{jk}} = (\theta_{j\ell}, \ell \in G_{jk})^T, \quad k = 0, 1, \ldots$$

where $G_{j0} = \{1, \ldots, \ell_{j0}\}$ of size $|G_{j0}| = \ell_{j0}$ and $G_{jk} = \{\ell_{jk-1} + 1, \ldots, \ell_{jk}\}$ of size $|G_{jk}| = \ell_{jk} - \ell_{jk-1} \leq (2^k/w_{nj})^{2\tau_j}$ for $k \geq 1$. Define $\tilde{\theta}_j$, a rescaled version of $\theta_j$, by

$$\tilde{\theta}_{j,G_{jk}} = (\tilde{\theta}_{j\ell}, \ell \in G_{jk}) = 2^k w_{nj}^{-1} \theta_{j,G_{jk}}.$$ 

It follows directly from (63) and (64) that

$$\|\tilde{\theta}_{j,G_{j0}}\|_2 \leq 1, \quad \|\tilde{\theta}_{j,G_{jk}}\|_{r_j} \leq (2^k/w_{nj})^{\ell_{j0}^{-1}(2\tau_j)} \leq 1 \forall k \geq 1, \forall f_j \in B_j.$$ 

Let $U_{jk} = \{u_{j,G_{jk}}(X_1), \ldots, u_{j,G_{jk}}(X_n)\}^T \in \mathbb{R}^{n \times |G_{jk}|}$. We have

$$\sup_{f_j \in B_j, f_{j'} \in B_{j'}} \left| \langle f_j, f_{j'} \rangle_n - \langle f_j, f_{j'} \rangle L_2 \right| = \sup_{f_j \in B_j, f_{j'} \in B_{j'}} \left| \left( \sum_{k=0}^\infty \sum_{\ell=0}^\infty \tilde{\theta}_{j,G_{jk}}^T \left[ U_{jk}^T U_{j',\ell} / n - EU_{jk}^T U_{j',\ell} / n \right] \tilde{\theta}_{j',G_{j'}} \right) \right| \leq \max_{\|\tilde{\theta}_j\|_2, \|\tilde{\theta}_{j'}\|_2} \left| \sum_{k=0}^\infty \sum_{\ell=0}^\infty \tilde{\theta}_{j,G_{jk}}^T \left[ U_{jk}^T U_{j',\ell} / n - EU_{jk}^T U_{j',\ell} / n \right] \tilde{\theta}_{j',G_{j'}} / \left(2^k w_{nj}^{10} w_{nj'}^{10} / \sqrt{5} \right) \right|.$$ 

(S42)

Let $a_k = 1/\{(k+1)(k+2)\}$. By (62), $\sup_{x \in \Omega} \|u_{j,G_{jk}}(x)\|^2 \leq \sup_{x \in \Omega} \sum_{k=0}^{\ell_{jk}} u_{j\ell}^2(x) \leq C_{5.1} \ell_{jk}$ for $k \geq 0$. By (65), $\|E U_{jk}^T U_{j',\ell} / n\|_S \leq C_{5.3}$. Because $|G_{jk}| \leq \ell_{jk}$, it follows from Lemma 9 that

$$\|U_{jk}^T U_{j',\ell} / n - EU_{jk}^T U_{j',\ell} / n\|_S \leq \left\{ \log(\ell_{jk} + 2) - \log(a_k a_\ell) + t \right\} 2 C_{5.1} C_{5.3} (\ell_{jk} + 2) / n + \left\{ \log(\ell_{jk} + 2) - \log(a_k a_\ell) + t \right\} (4/3) C_{5.1} \sqrt{\ell_{jk} \ell_{j'} / n}.$$  

(S43)

with probability at least $1 - a_k a_\ell e^{-t}$ for any fixed $k \geq 0$ and $\ell \geq 0$. By the union bound and the fact that $\sum_{k=0}^\infty a_k = 1$, inequality (S43) holds simultaneously for all $k \geq 0$ and $\ell \geq 0$ with probability at least $1 - e^{-t}$. Because $\ell_{jk} = [(2^k/w_{nj})^{2\tau_j}]$, we rewrite (S43) as

$$\|U_{jk}^T U_{j',\ell} / n - EU_{jk}^T U_{j',\ell} / n\|_S \leq C_{5.4} \left[ \left( 2^{\tau_j} w_{nj}^{-\tau_j} + 2^{\tau_j} w_{nj'}^{-\tau_j} \right)^{\log \left( \left( k + \ell + \log(w_{nj}^{-\tau_j} + w_{nj'}^{-\tau_j}) + t \right) / n \right)} + \left( 2^{\tau_j} w_{nj}^{-\tau_j} + w_{nj'}^{-\tau_j} \right)^{\log \left( \left( 2^{\tau_j} w_{nj}^{-\tau_j} + w_{nj'}^{-\tau_j} \right) / n \right)} \right].$$  

(S44)
where \( C_{5,4} \) is a constant depending only on \( \{ C_{5,1}, C_{5,3} \} \). For any \( \alpha \geq 0, \sum_{k=0}^{\infty} k^{\alpha} 2^{-k(1-\tau_j)} \leq C_{\alpha} \mu_j^{\alpha+1} \), where \( C_{\alpha} \) is a numerical constant and \( \mu_j = 1/(1-\tau_j) \). Using this fact and inserting (S44) into (S42) yields the desired result.

Finally, the following result concludes the proof of Theorem 5.

**Lemma 11.** In the setting of Theorem 5, let

\[
\phi_n = C_{5,2} C_{5,4} \left\{ \max_j \frac{\sqrt{2 \log(np/\epsilon')}}{\lambda_{nj}} \mu_j \bar{r}_{nj}, \frac{2 \log(np/\epsilon') \mu_j^2 \bar{r}_{nj}^2}{\lambda_{nj}^2} \right\},
\]

where \( \bar{r}_{nj} = n^{-1/2} w_{nj}^{-\tau_j} \), \( \mu_j = 1/(1-\tau_j)^{-1} \), and \( C_{5,4} \) is a constant depending only on \( \{ C_{5,1}, C_{5,3} \} \) as in Lemma 10. Then

\[
P \left\{ \sup_{g \in \mathcal{G}} \left| \frac{\|g\|^2_n - \|g\|^2_Q}{R_n^2(g)} \right| \geq \phi_n \right\} \leq \epsilon'^2.
\]

**Proof.** Recall that \( \ell_{jk} = \lfloor (2^k/w_{nj})^{2\tau_j} \rfloor \). For \( g_j = \sum_{\ell=1}^{\infty} \theta_{j,\ell} u_{j,\ell} \), define \( r_{nj}(g_j) \) by

\[
r_{nj}^2(g_j) = \left( \sum_{\ell=1}^{\ell_{j0}} \theta_{j,\ell}^2 / w_{nj}^2 \right) \vee \left( \max_{k \geq 1} \sum_{\ell_{j,k-1} < \ell \leq \ell_{jk}} \theta_{j,\ell}^2 \ell_{j,\ell}^{1/\tau_j} \right).
\]

Let \( f_j = g_j / r_{nj}(g_j) \) and \( \mu_j = 1/(1-\tau_j) \). Then \( f_j \in \mathcal{B}_2 \) as in Lemma 10 and

\[
\left| \|g\|^2_n - \|g\|^2_Q \right| \leq \sum_{j=1}^{p} \sum_{j'=1}^{p} \left| \langle g_j, g_{j'} \rangle_n - \langle g_j, g_{j'} \rangle_Q \right|
= \sum_{j=1}^{p} \sum_{j'=1}^{p} r_{nj}(g_j) r_{nj'}(g_{j'}) \left| \langle f_j, f_{j'} \rangle_n - \langle f_j, f_{j'} \rangle_Q \right|.
\]

Because \( \sum_{j=1}^{p} w_{nj} \lambda_{nj} r_{nj}(g_j) \leq \sum_{j=1}^{p} C_{5,2}^{1/2} \lambda_{nj} (w_{nj} \|g_j\|_{F,j} + \|g_j\|_Q) = C_{5,2}^{1/2} R_n^2(g) \) by (63),

\[
\left\{ \sup_{g \in \mathcal{G}} \left| \frac{\|g\|^2_n - \|g\|^2_Q}{R_n^2(g)} \right| \geq \phi_n \right\}
\subset \bigcup_{j,j'} \left\{ \sup_{f_j \in B_j, f_{j'} \in B_{j'}} \left| \langle f_j, f_{j'} \rangle_n - \langle f_j, f_{j'} \rangle_Q \right| > C_{5,2}^{-1/2} \phi_n w_{nj} \lambda_{nj} w_{nj'} \lambda_{nj'} \right\}.
\]

By Lemma 10 with \( t = \log(p^2/\epsilon'^2) \) and \( e^{\epsilon'^2} + 2 w_{nj}^{-\tau_j} \leq n \), we have

\[
\sup_{f_j \in B_j, f_{j'} \in B_{j'}} \left| \langle f_j, f_{j'} \rangle_n - \langle f_j, f_{j'} \rangle_Q \right|
\leq C_{5,4} w_{nj} w_{nj'} \left\{ (\mu_j w_{nj}^{-\tau_j} + \mu_j' w_{nj'}^{-\tau_j'}) \sqrt{\{ \mu_j + \mu_j' + \log(\sum_{j=1}^{p} \lambda_{nj} \|w_{nj}^{-\tau_j} + w_{nj'}^{-\tau_j'}\|_n)} + \log(p^2/\epsilon'^2)\} / n \right.
+ \{ \mu_j + \mu_j' + \log(\sum_{j=1}^{p} \lambda_{nj} \|w_{nj}^{-\tau_j} + w_{nj'}^{-\tau_j'}\|_n)} \left( \mu_j w_{nj}^{-\tau_j} \mu_j' w_{nj'}^{-\tau_j'} \right) / n \right\}
\leq C_{5,4} w_{nj} w_{nj'} \left\{ (\mu_j w_{nj}^{-\tau_j} + \mu_j' w_{nj'}^{-\tau_j'}) \sqrt{2 \log(np/\epsilon')} / n + 2 \log(p/\epsilon')(\mu_j w_{nj}^{-\tau_j})(\mu_j' w_{nj'}^{-\tau_j'}) / n \right\},
\]

23
with probability at least $1 - \epsilon^2/p^2$. By the definition of $\phi_n$, we have

$$P \left\{ \sup_{f_j \in B_j, f_j' \in B_j'} |\langle f_j, f_j'\rangle_n - \langle f_j, f_j'\rangle_Q| \leq C_{5,2}^{-1} \phi_n w_n j \cdot \lambda_n j \cdot \lambda_n j' \right\} \geq 1 - \frac{\epsilon^2}{p^2}.$$ 

The conclusion follows from the union bound using (S45).

\[ \square \]

## S2 Technical tools

### S2.1 Sub-Gaussian maximal inequalities

The following maximal inequality can be obtained from van de Geer (2000, Corollary 8.3), or directly derived using Dudley’s inequality for sub-Gaussian variables and Chernoff’s tail bound (see Proposition 9.2, Bellec et al. 2016).

**Lemma 12.** For $\delta > 0$, let $F_1$ be a functional class such that $\sup_{f_1 \in F_1} \|f_1\|_n \leq \delta$, and

$$\psi_n(\delta, F_1) \geq \int_0^\delta H^{1/2}(u, F_1, \|\cdot\|_n) \, du. \quad (S46)$$

Let $(\varepsilon_1, \ldots, \varepsilon_n)$ be independent sub-Gaussian variables under Assumption 1. Then for any $t > 0$,

$$P \left\{ \sup_{f_1 \in F_1} |\langle \varepsilon, f_1\rangle_n| / C_1 > n^{-1/2} \psi_n(\delta, F_1) + \delta \sqrt{1/n} \right\} \leq \exp(-t),$$

where $C_1 = C_1(D_0, D_1) > 0$ is a constant, depending only on $(D_0, D_1)$.

### S2.2 Dudley and Talagrand inequalities

The following inequalities are due to Dudley (1967) and Talagrand (1996).

**Lemma 13.** For $\delta > 0$, let $F_1$ be a functional class such that $\sup_{f_1 \in F_1} \|f_1\|_n \leq \delta$ and (S46) holds. Let $(\sigma_1, \ldots, \sigma_n)$ be independent Rademacher variables, that is, $P(\sigma_i = 1) = P(\sigma_i = -1) = 1/2$. Then for a universal constant $C_2 > 0$,

$$E \left\{ \sup_{f_1 \in F_1} |\langle \sigma, f_1\rangle_n| / C_2 \right\} \leq n^{-1/2} \psi(\delta, F_1).$$

**Lemma 14.** For $\delta > 0$ and $b > 0$, let $(X_1, \ldots, X_n)$ be independent variables, and $F$ be a functional class such that $\sup_{f \in F} \|f\|_Q \leq \delta$ and $\sup_{f \in F} \|f\|_\infty \leq b$. Define

$$Z_n = \sup_{f \in F} \left| \frac{1}{n} \sum_{i=1}^n \{ f(X_i) - E f(X_i) \} \right|.$$
Then for a universal constant $C_3 > 0$, we have

$$P \left\{ \frac{Z_n}{C_3} > E(Z_n) + \delta \sqrt{\frac{t}{n} + b \frac{t}{n}} \right\} \leq \exp(-t), \quad t > 0.$$  

### S2.3 Non-commutative Bernstein inequality

We state the non-commutative Bernstein inequality (Troop, 2011) as follows.

**Lemma 15.** Let $\{M_i : i = 1, \ldots, n\}$ be independent random matrices in $\mathbb{R}^{d_1 \times d_2}$ such that $E(M_i) = 0$ and $P\{\|M_i\| \leq s_0\} = 1$, $i = 1, \ldots, n$, for a constant $s_0 > 0$, where $\| \cdot \|_S$ denotes the spectrum norm of a matrix. Let $\Sigma_{\text{col}} = \sum_{i=1}^n E(M_iM_i^T)/n$ and $\Sigma_{\text{row}} = \sum_{i=1}^n E(M_i^TM_i)/n$. Then, for all $t > 0$,

$$P\left\{ \frac{1}{n} \sum_{i=1}^n M_i > t \right\} \leq (d_1 + d_2) \exp \left( \frac{-n t^2 / 2}{\| \Sigma_{\text{col}} \|_S \sqrt{\| \Sigma_{\text{row}} \|_S + s_0 t / 3}} \right).$$  

Consequently, for all $t > 0$,

$$P\left\{ \frac{1}{n} \sum_{i=1}^n M_i > \sqrt{\| \Sigma_{\text{col}} \|_S \sqrt{\| \Sigma_{\text{row}} \|_S}} \sqrt{2t / n} + (s_0 / 3)2t / n \right\} \leq (d_1 + d_2)e^{-t}.$$  

### S2.4 Convergence of empirical norms

For $\delta > 0$ and $b > 0$, let $\mathcal{F}_1$ be a functional class such that

$$\sup_{f_1 \in \mathcal{F}_1} \| f_1 \|_Q \leq \delta, \quad \sup_{f_1 \in \mathcal{F}_1} \| f_1 \|_\infty \leq b,$$

and let $\psi_{n,\infty}(\cdot, \mathcal{F}_1)$ be an upper envelope of the entropy integral:

$$\psi_{n,\infty}(z, \mathcal{F}_1) \geq \int_0^z H^*(u/2, \mathcal{F}_1, \| \cdot \|_{n,\infty}) \, du, \quad z > 0,$$

where $H^*(u, \mathcal{F}_1, \| \cdot \|_{n,\infty}) = \sup_{(X_1^{(1)}, \ldots, X_n^{(1)})} H(u, \mathcal{F}_1, \| \cdot \|_{n,\infty})$. Let $\delta = \sup_{f_1 \in \mathcal{F}_1} \| f_1 \|_n$. The following result can be obtained from Guedon et al. (2007) and, in its present form, van de Geer (2014), Theorem 2.1.

**Lemma 16.** For the universal constant $C_2$ in Lemma 13, we have

$$E \left\{ \sup_{f_1 \in \mathcal{F}_1} \| f_1 \|_n^2 - \| f_1 \|_Q^2 \right\} \leq \frac{2\delta C_2 \psi_{n,\infty}(b, \mathcal{F}_1)}{\sqrt{n}} + \frac{4C_2^2 \psi^2_{n,\infty}(b, \mathcal{F}_1)}{n}.$$  

Moreover, we have

$$\sqrt{E(\delta^2)} \leq \delta + \frac{2C_2 \psi_{n,\infty}(b, \mathcal{F}_1)}{\sqrt{n}}.$$
S2.5 Metric entropies

For \( r \geq 1 \) and \( m > 0 \) (possibly non-integral), let \( \mathcal{W}_r^m = \{ f : \| f \|_{L_r} + \| f^{(m)} \|_{L_r} \leq 1 \} \). The following result is taken from Theorem 5.2, Birman & Solomjak (1967).

Lemma 17. If \( rm > 1 \) and \( 1 \leq q \leq \infty \), then

\[
H(u, \mathcal{W}_r^m, \cdot \cdot \cdot L_q) \leq B_1 u^{-1/m}, \quad u > 0,
\]

where \( B_1 = B_1(m, r) > 0 \) is a constant depending only on \( (m, r) \). If \( rm \leq 1 \) and \( 1 \leq q < r/(1 - rm) \), then

\[
H(u, \mathcal{W}_r^m, \cdot \cdot \cdot L_q) \leq B_2 u^{-1/m}, \quad u > 0,
\]

where \( B_2 = B_2(m, r, q) > 0 \) is a constant depending only on \( (m, r, q) \).

For \( m \geq 1 \), let \( \mathcal{V}_r^m = \{ f : \| f \|_{L_1} + \text{TV}(f^{(m-1)}) \leq 1 \} \). The following result can be obtained from Theorem 15.6.1, Lorentz et al. (1996), on the metric entropy of the ball \( \{ f : \| f \|_{L_r} + [f]_{\text{Lip}(m, L_r)} \leq 1 \} \), where \([f]_{\text{Lip}(m, L_r)}\) is a semi-norm in the Lipschitz space Lip\((m, L_r)\). By Theorem 9.9.3, DeVore & Lorentz (1993), the space Lip\((m, L_1)\) is equivalent to \( \mathcal{V}_r^m \), with the semi-norm \([f]_{\text{Lip}(m, L_r)}\) equal to \( \text{TV}(f) \), up to suitable modification of function values at (countable) discontinuity points. However, it should be noted that the entropy of \( \mathcal{V}_r^1 \) endowed with the norm \( \cdot \cdot \cdot L_\infty \) is infinite.

Lemma 18. If \( m \geq 2 \) and \( 1 \leq q \leq \infty \), then

\[
H(u, \mathcal{V}_r^m, \cdot \cdot \cdot L_q) \leq B_3 u^{-1/m}, \quad u > 0,
\]

where \( B_3 = B_3(m) > 0 \) is a constant depending only on \( m \). If \( 1 \leq q < \infty \), then

\[
H(u, \mathcal{V}_r^1, \cdot \cdot \cdot L_q) \leq B_4 u^{-1}, \quad u > 0,
\]

where \( B_4 = B_4(r) > 0 \) is a constant depending only on \( r \).

By the continuity of functions in \( \mathcal{W}_r^m \) for \( m \geq 1 \) and \( \mathcal{V}_r^m \) for \( m \geq 2 \), the maximum entropies of these spaces in \( \cdot \cdot \cdot \cdot \cdot L_\infty \) and \( \cdot \cdot \cdot \cdot \cdot L_\infty \) norms over all possible design points can be derived from Lemmas 17 and 18.

Lemma 19. If \( rm > 1 \), then for \( B_1 = B_1(m, r) \),

\[
H^*(u, \mathcal{W}_r^m, \cdot \cdot \cdot n) \leq H^*(u, \mathcal{W}_r^m, \cdot \cdot \cdot n) \leq B_1 u^{-1/m}, \quad u > 0,
\]
and hence (S46) and (S49) hold with \( \psi_n(z, \mathcal{W}_r^m) \preceq \psi_{n, \infty}(z, \mathcal{W}_r^m) \preceq z^{1-1/(2m)} \). If \( m \geq 2 \), then for \( B_3 = B_3(m) \),

\[ H^*(u, \mathcal{V}_r^m, \| \cdot \|_n) \leq H^*(u, \mathcal{V}_r^m, \| \cdot \|_{n, \infty}) \leq B_3 u^{-1/m}, \quad u > 0, \]

and hence (S46) and (S49) hold with \( \psi_n(z, \mathcal{V}_r^m) \preceq \psi_{n, \infty}(z, \mathcal{V}_r^m) \preceq z^{1-1/(2m)} \).

The maximum entropies of \( \mathcal{V}_1 \) over all possible design points can be obtained from Section 5, Mammen (1991) for the norm \( \| \cdot \|_n \) and Lemma 2.2, van de Geer (2000) for the norm \( \| \cdot \|_{n, \infty} \). In fact, the proof of van de Geer shows that for \( \mathcal{F} \) the class of nondecreasing functions \( f : [0, 1] \to [0, 1] \), \( H^*(u, \mathcal{F}, \| \cdot \|_{n, \infty}) \leq n \log(n + u^{-1}) \) if \( u \leq n^{-1} \) or \( \leq u^{-1} \log(n + u^{-1}) \) if \( u > n^{-1} \). But if \( u \leq n^{-1} \), then \( n \log(n + u^{-1}) \leq n \log n + u^{-1} \log(n + u^{-1}) \leq (1 + \log n)u^{-1} \). If \( u > n^{-1} \), then \( u^{-1} \log(n + u^{-1}) \leq u^{-1} \log(2n) \). Combining the two cases gives the stated result about \( H^*(u, \mathcal{V}_1^1, \| \cdot \|_{n, \infty}) \), because each function in \( \mathcal{V}_1^1 \) can be expressed as a difference two nondecreasing functions.

**Lemma 20.** For a universal constant \( B_5 > 0 \), we have

\[ H^*(u, \mathcal{W}_1^m, \| \cdot \|_n) \leq H^*(u, \mathcal{V}_1^1, \| \cdot \|_n) \leq B_5 u^{-1}, \quad u > 0, \]

and hence (S46) holds with \( \psi_n(z, \mathcal{W}_1^m) \preceq \psi_n(z, \mathcal{V}_1^1) \preceq z^{1/2} \). Moreover, for a universal constant \( B_6 > 0 \), we have

\[ H^*(u, \mathcal{W}_1^m, \| \cdot \|_{n, \infty}) \leq H^*(u, \mathcal{V}_1^1, \| \cdot \|_{n, \infty}) \leq B_6 \frac{1 + \log n}{u}, \quad u > 0, \]

and hence (S49) holds with \( \psi_{n, \infty}(z, \mathcal{W}_1^m) \preceq \psi_{n, \infty}(z, \mathcal{V}_1^1) \preceq (1 + \log n)^{1/2}(z/2)^{1/2} \).

### S2.6 Interpolation inequalities

The following inequality (S50) can be derived from the Gagliardo-Nirenberg inequality for Sobolev spaces (Theorem 1, Nirenberg 1966). Inequality (S51) can be shown by approximating \( f \in \mathcal{V}_m^m \) by functions in \( \mathcal{W}_1^m \).

**Lemma 21.** For \( r \geq 1 \) and \( m \geq 1 \), we have for any \( f \in \mathcal{W}_r^m \),

\[
\|f\|_{\infty} \leq (C_4/2) \left\{ \|f^{(m)}\|_{L_r} + \|f\|_{L_2} \right\}^\tau \|f\|_{L_2}^{1-\tau}, \tag{S50}
\]

where \( \tau = (2m + 1 - 2/r)^{-1} \leq 1 \) and \( C_4 = C_4(m, r) \geq 1 \) is a constant depending only on \( (m, r) \).

In addition, we have for any \( f \in \mathcal{V}_m^m \),

\[
\|f\|_{\infty} \leq (C_4/2) \left\{ \text{TV}(f^{(m-1)}) + \|f\|_{L_2} \right\}^\tau \|f\|_{L_2}^{1-\tau}. \tag{S51}
\]
From this result, $\|f\|_{\infty}$ can be bounded in terms of $\|f\|_{L_2}$ and $\|f^{(m)}\|_{L_r}$ or $\text{TV}(f^{(m-1)})$ in a convenient manner. For $f \in \mathcal{W}_r^m$ and $0 < \delta \leq 1$, if $\|f\|_{L_2} \leq \delta$ and $\|f^{(m)}\|_{L_r} \leq 1$, then $\|f\|_{\infty} \leq C_4 \delta^{1-1/(2m+1-2/r)}$. Similarly, for $f \in \mathcal{V}^m$ and $0 < \delta \leq 1$, if $\|f\|_{L_2} \leq \delta$ and $\text{TV}(f^{(m-1)}) \leq 1$, then $\|f\|_{\infty} \leq C_4 \delta^{1-1/(2m-1)}$.

References


